

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT

Ranel Tammela

Melanoomi antigeene kandvate viiruslaadsete partiklite proteoomianalüüs

Bakalaureusetöö

Juhendajad *PhD* Reet Kurg,

PhD Aare Aabroi

Tartu 2016

Sisukord

Kasutatud lühendid.....	4
Sissejuhatus.....	6
1. Kirjanduse ülevaade.....	8
1.1. Proteoomianalüüs märgisevaba kvantifitseerimise abil.....	8
1.2. Hierarhiline klasteranalüüs.....	10
1.3. Kvantiil-kvantiil graafik.....	13
1.4. <i>Residuals vs fitted</i>	13
1.5. <i>Residuals vs leverage</i>	15
1.6. Korrelatsioonikordaja seos korrigeeritud determinatsioonikordajaga.....	16
1.7. Ekstratsellulaarsete partiklite genereerimine MLV gag valguga abil.....	16
2. Eksperimentaalne osa.....	18
2.1. Eesmärgid.....	18
2.2. Materjal ja meetodika.....	18
2.2.1. VLP-de saamine ja mass-spektromeetriline analüüsimine.....	18
2.2.2. Regressioonanalüüs R-studio programmiga.....	19
2.2.3. Hierarhiline klasterdamine proteoomianalüüsi mustrite leidmiseks.....	19
2.3. Tulemused ja arutelu.....	20

2.3.1. Proteoomide võrdlus, kasutades Q-Q graafikute meetodit.....	20
2.3.2. Regressioonigraafikute analüüs.....	22
2.3.3. <i>Residuals vs fitted</i> graafikute analüüs.....	23
2.3.4. <i>Residuals vs leverage</i> graafikute analüüs.....	24
2.3.5. Hierarhilise klasterdamise tulemused.....	26
2.4. Arutelu.....	27
Kokkuvõte.....	31
Proteomic analysis of virus-like particles carrying melanoma antigens.....	33
Tänuõnad.....	35
Kasutatud kirjandus.....	36
Kasutatud veebiaadressid.....	41
Lisa 1.....	42
Lisa 2.....	57

Kasutatud lühendid

BSA – veise albumiini seerum (*Bovine serum albumin*)

CD146 – diferentseerimisklaster 146 (*Cluster of differentiation 146*)

CMV – tsütomegaloviirus (*Cytomegalovirus*)

COP5 – hiire fibroplastirakud

CSV – komaga eraldatud väärtused (*Comma-separated values*)

DLS – dünaamiline valguse hajuvus (*Dynamic Light Scattering*)

DTT – ditiotreitool (*Dithiothreitol*)

EBNA – Epstein-Barri viiruse tuumaantigeen (*Epstein-Barr virus-encoded nuclear antigen*)

FCS – veiselote seerum (*Fetal calf serum*)

hEF1a – inimese elongatsioonifaktor 1-Alpha (*human Elongation Factor 1-Alpha*)

IMDM – Iscove'i modifitseeritud Dulbecco sööde (*Iscove's Modified Dulbecco's Medium*)

iTRAQ – isobaarsed märgistamised relatiivseks ja absoluutseks kvantifitseerimiseks (*Isobaric tags for relative and absolute quantification*)

LC-MS – vedelikkromatograafia-mass-spektromeetria (*Liquid chromatography-mass spectrometry*)

LFQ – märgisevaba kvantifitseerimine (*Label-Free Quantification*)

MAGEA4 – melanoomi assiotsieerunud antigeen 4 (*Melanoma-associated antigen 4*)

MAGEA10 – melanoomi assiotsieerunud antigeen 10 (*Melanoma-associated antigen 10*)

MART1 – melanoomi T-rakkudega 1 tuvastatud antigeen (*Melanoma antigen recognized by T cells 1*)

MCAM – melanoomirakkude adhesioonimolekulid (*Melanoma cell adhesion molecule*)

MLV – hiire leukeemiaviirus (*Murine leukemia virus*)

MS² ehk MS/MS – tandem mass-spektromeetria (*Second stage of mass spectrometry*)

MUC18 – rakupinna glükoproteiin (*Cell surface glycoprotein MUC18*)

PBS – fosfaatpuhvri lisandiga keedusoolalahus (*Phosphate buffered saline*)

SILAC – stabiilsete isotoopide aminohapete rakukultuuris märgistamine (*Stable isotope labeling using amino acids in cell culture*)

TMT – tandeemsed massi märgised (*Tandem mass tags*)

TRP1 – fosfo-ribosüül-antraniilaadi isomeraas (*Phosphoribosylanthranilate isomerase*)

VLP – viiruslaadne partikkel (*Virus-like particle*)

Q-Q – kvantiil-kvantiil (*Quantile-Quantile*)

XIC – valitud iooni seire kromatogramm (*Selected-ion monitoring chromatogram*)

Sissejuhatus

Suurem osa rakutüüpidest vabastab ekstratsellulaarseid vesiikuleid – membraaniga ümbritsetud valkude, nukleiinhapete ja teiste biomolekulide kandjaid, mis osalevad nende molekulide transpordis rakkude ja kudede vahel, olles kaasatud rakkudevahelisse kommunikatsiooni, koagulatsiooni, tuumorigeneesi ja immuunraktsiooni protsessi. Terminit ekstratsellulaarne vesiikul kasutatakse mis tahes sekreteeritud vesiikulite – nii eksosoomide, mikrovesiikulite kui apoptootiliste kehade kohta. Mikrovesiikulid, mida mõnikord nimetatakse ka mikroosakesteks, on plasmamembraanist saadud 100–1000 nm läbimõõduga osakesed, mis tekivad plasmamembraani väljapoole pungumisel. Sarnaselt ekstratsellulaarsete mikrovesiikulite pungumisega punguvad ka retroviirused, kus viiruse gag valgumolekul indutseerib spontaanset mikrovesiikulite või viiruslaadsete osakeste teket imetajate rakkudes. Viiruslaadseid partikleid saab kasutada, esitamaks immuunsüsteemile võõrepitoope. Gag põhise viiruse laadseid partikleid kasutatakse vaktsiinidena, sest need stimuleerivad humoraalset ja rakulist immuunvastust. (Kurg jt, 2016.)

Melanoomiga seotud antigeenid (MAGE), mis indutseerivad tsütotoksiliste T-lümfotsüütide teket, kuuluvad vähi-testise antigeenide perekonda ja avastati melanoomirakkude uurimisel. Valke MAGEA10, MAGEA4, MART1, TRP1 ja MCAM ekspresseeritakse paljudes vähitüvedes ja seetõttu peetakse neid potentsiaalseks sihtmärgiks vähktõvevaktsiinide ja immunoteraapia loomisel. Uuringus „Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens” (Kurg jt, 2016) genereeriti viiruslaadseid partikleid MLV gag valku kasutades. Melanoomi antigeenid MAGEA4, MAGEA10, MART1, TRP1 ja MCAM koekspresseeriti koos MLV gag valguga hiire fibroblastides ja mõõdeti valkude suhtelist arvukust saadud rakuvälise partiklite sees.

Käesoleva bakalaaurusetöö eesmärk on leida eelnevalt mainitud valkude seosed teiste valkude suhtelise arvukusega mikrovesiikulites, kasutades LFQ intensiivsuse analüüsi pakettidega R-studio ja SPSS. Analüüsimiseks on kasutatud andmeid, mis on saadud

uurimusest „Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens” (Kurg jt, 2016).

Töö on koostatud Tartu Ülikooli Tehnoloogiainstituudis.

1. Kirjanduse ülevaade

1.1. Proteoomianalüüs märgisevaba kvantifitseerimise abil

Mass-spektromeetriast on saanud üks proteoomika olulisi analüüsimismeetodeid valkude identifitseerimisel ja kvantifitseerimisel. Hiljutised uuringud on näidanud, et mass-spektromeetria annab proteoomika eksperimentides lisaks valkude tuvastamisele nende kohta ka olulist kvantitatiivset informatsiooni – informatsiooni nende absoluutse või suhtelise koguse kohta. Proteoomika valdkonnas on kvantifitseerimisel kasutusele võetud mitmeid märgistamismeetodeid: isotoopne, metaboolne ja isobaarne märgistus. Märgisevaba kvantifitseerimise korral seevastu ei kasutata märgistamist mitte kummaski proovis ning see meetod on teistest proteoomikas kasutusel olevatest meetoditest kiirem ja tundlikum. (Nahnsen jt, 2013.) Võrreldes näiteks meetoditega nagu SILAC, iTRAQ ja TMT, on märgisevabal kvantifitseerimisel suurem dünaamiline ulatus ja kõrgem identifitseerimissuutlikkus. Samas on märgisevaba kvantifitseerimise kvantitatiivne täpsus väiksem kui märgistamisel põhinevatel meetoditel. (Megger jt, 2013.)

Mass-spektromeetrial põhinev märgisevaba kvantifitseerimine jaguneb kahte kategooriasse. Esimese puhul tuginetakse peamiselt peptiidide intensiivsusele, mõõtes ioonide kromatograafilist intensiivsuse muutust (peptiidide piigipindalad ja -kõrgused), seda nimetatakse märgisevabaks vedelikkromatograafia-mass-spektromeetria ehk LC-MS meetodiks, millega tuvastatakse aineid massi ja laengu suhte järgi. (Clough jt, 2012.)

Niisugune märgisevaba signaalitugevusel põhinev valgu kvantifitseerimine juhusliku ehk *shotgun*-sekveneerimise meetodiga on paljulubav alternatiiv stabiilsete isotoopide märgistamisele ekspresseeritud valkude hindamisel kompleksetes bioloogilistes proovides. See lihtne ja tulus meetod võimaldab statistilise analüüsi kaasamise korral tuvastada ka väikseid bioloogiliselt mõtestatud muutusi. Mitmed uuringud on näidanud, et valitud peptiidi-ioonide ekstraheeritud ionkromatogrammide (XIC) korreleeruvad hästi valkude suhtelise arvukusega bioloogiliste proovide kogumis.

Teine märgisevaba kvantifitseerimise meetod, mida nimetatakse spektraalseks loendamiseks, võrdleb igale valgule määratud MS/MS spektreid, mõõtes niiviisi valkude suhtelist arvukust. MS/MS ehk tandem-mass-spektromeetria on kahest protsessist koosnev fragmentatsioon, mille esialgne seleksioon põhineb aparatuuri võimel viia läbi vaakumi kaks või enam ioniseeritud molekulaarset partiklit, millel on sama massi ja laengu suhe. Teises staadiumis eraldatakse molekulid üksteisest ning hiljem detekteeritakse ja loendatakse. Seda staadiumi nimetatakse märgisevabaks MS² kvantifitseerimiseks. Spektri fragmentatsiooni järgselt on MS² kvantifitseerimisel olulisim analüüsimisprotseduur kõige sobivama matemaatilise algoritmi valimine. Selle meetodi rakendamine nõuab suuremat arvutusvõimsust ja niisuguseid algoritme, mis suudavad joondada kromatograafilisi piike ja mõõta iooni intensiivsust, analüüsivaks valkude suhtelise arvukuse muutust bioloogiliste proovide kogumis. (Everitt jt, 2011.)

Märgisevabale kvantifitseerimisele järgneva statistilise analüüsiga on võimalik tuvastada ka väikseid olulisi muutusi, mis on bioloogiliselt mõtestatud. Mitmed uuringud on näidanud, et nii spektraalne loendamine kui valitud peptiidi ionide kromatograafiline ekstraheerimine korreleeruvad hästi valgu kompleksse suhtelise arvukusega bioloogilistes proovides. (Wang jt, 2008.)

Sõltumata sellest, millist märgisevaba kvantifikatsiooni meetodit proteoomikas kasutatakse, hõlmab protseduur alati kindlaid etappe, mille hulka kuuluvad proovi ettevalmistamine ehk valkude ekstraheerimine, redutseerimine, alküülimine ja lagundamine, seejärel proovi lahutamine vedelikkromatograafiaga ja selle analüüsimine, lõpuks valkude identifitseerimine, kvantifitseerimine ja analüüs. Valkude märgistamise meetod sisaldab proovide omavahelist kombineerimist pärast märgistamise lõpetamist.

Märgistamise meetodist paindlikuma märgisevaba kvantifikatsiooni meetodi puhul seevastu valmistatakse iga proov eraldi, niisiis ei ole ka proovi suurus ühelgi moel piiratud. (Pocsfalvi jt, 2015).

1.2. Hierarhiline klasteranalüüs

Klasteranalüüsi eesmärk on liigitada objektid organiseeritud süsteemi, kus igas rühmas on sarnaste omadustega liikmed. Niisugused rühmad on suheliselt homogeenised iseenese suhtes ja heterogeensed omavahel. Oluline on märkida, et klasterdamise protseduuri eesmärk pole mitte gruppide erinevuste välja selgitamine, vaid juhtumite või objektide arvu vähendamine neid homogeensetesse klastridesse liigitades.

Klasteranalüüs võib olla hierarhiline, jaotav (*partitional*) või alamruumiline (*subspace*). Proteoomianalüüsis on hierarhiline klasterdamine eelistatavam, kuna sellega on võimalik analüüsi teostada erinevatel tasanditel ja visualiseerida tulemusi.

Hierarhilist klasterdamist võib jagada alt üles ehk aglomeratiivseks ja ülalt alla ehk eristavaks meetodiks. Alt üles ehk aglomeratiivset meetodit kasutatakse rohkem. Sel juhul vaadeldakse alguses iga valku ühe klastrina ja ühendatakse seejärel kaks klastrit rekursiivselt omavahel nii, et nende distants oleks minimaalne, sama tegevust jätkatakse uute tekkinud klastritega, kuni kogu andmestik saab liidetud üheks klastriks. (Shiga, Mamitsuka, 2011.)

Reeglina on niisugune tegevus jaotatud neljaks sammuks. Esimese sammu puhul arvestatakse, et iga valk vastab ühele klastrile. Teise sammuna tuleb leida kõikide võimalike paaride vahekaugus, kolmanda sammuna ühendada leitud lähedased klastrid üheks klastriks. Neljanda sammuna tuleb kontrollida, kas klasterdamine on õigesti tehtud või peab midagi muutma. Iga hierarhilise klasterdamise käigus teostatud ühendamine, millega saadakse uus, sarnastest objektidest grupeeritud klaster, on pöördumatu – objekte ei saa enam teise klastrisse määrata. Saadud klastritega liidetakse hiljem ülejäänud objektid. (Zhang, Zhang, 2006.)

Sarnaste objektide kauguse leidmiseks on klasteranalüüsis neli võimalikku meetodit. Levinuim neist on Eukleidese kauguse meetod, mille puhul vahemaa leitakse valemiga:

$$D(g_1, g_2) = \sqrt{\sum_{j=1}^m (W_{1j} - W_{2j})^2} \quad (1)$$

Siin on g_1 ja g_2 kaks andmekogumi valku, m tähistab dimensiooninumbrit või tingimust. W_{1j} ja W_{2j} on g_1 ja g_2 valkude profiili väärtused.

See algoritm võimaldab leida vahemaa kahe objekti vahel kõikide muutujate puhul. Eukleidese ruudu distantse kuvatakse iga sammu puhul kõikide paaris olevate objektide vahe kalkuleeritud maatriksina. Igas etapis liidetakse omavahel kokku need paaris olevad objektid või klastrid, millel on kõige minimaalsem Eukleidese vahemaa. See muudab hierarhilise klasterdamise käigus toimuva rühmitamise pikaajaliseks protsessiks, kuna iga sammu korral peab arvutama uue maatriksi, et võtta arvesse hiljuti ühendatud klastrid.

Eukleidese kauguse meetodi kasutamine on lihtne, kui tegemist on ühe objektiga ühe klasteri kohta. Kui ühes või mõlemas klasteris on üle ühe objekti, tuleb kasutada mõnda muud seose leidmise meetodit, et arvutada täpne vahekaugus paaris olevate klasterite iga objekti puhul. Taas on eesmärgiks leida kaks klasterit, mis on teineteisele kõige lähemal, et need omavahel ühendada.

Kolme ja rohkema objektide arvuga kahe klasteri vahemaad määrava seose leidmiseks on mitmeid meetodeid. Neist ühe sideme seost nimetatakse ka lähima naabri seoseks või minimaalseks seoseks. See meetod määrab kauguse kahe klasteri vahel, leides kahe klasteri objektide minimaalse distantse.

Klastrite x ja y distantis leitakse järgmise valemiga:

$$D(x,y) = \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} \text{Min}\{d(x_i, y_j)\}$$

kus S_x ja S_y on klastrite suurused. Ühe sideme seose puuduseks on asjaolu, et see võib mõnikord põhjustada klastrite aheldatust, andes seetõttu vääralt lõpptulemuse. Täielik seostatus ehk kaugeima naabri ehk maksimaalse seose meetod sarnaneb ühe sideme seosega, kuid erineb viimatinimetatust selle poolest, et siin otsitakse klastripaari objektide vahelist suurimat distantssi. Klastrite x ja y distantis leitakse järgmise valemiga:

$$D(x,y) = \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} \text{Max}\{d(x_i, y_j)\}$$

Siin tekib probleem põhjusel, et äärealade objekte otsides jäetakse klastrite lähedasemad objektid kõrvale.

On olemas ka keskmise seostatuse ehk kaalumata keskmise seostatusega meetod, mis peaks andma klastrite vahelise kauguse kohta täpseima hinnangu. Selle puhul leitakse igas klastris keskmised objektid ja määratakse nendevaheline distantis. Klastrite x ja y distantisi leidmiseks kasutatakse järgmist valemit:

$$D(x,y) = \frac{1}{S_x S_y} \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} d(x_i, y_j)$$

(Hussain, Hazarika, 2011.)

Hierarhiline aglomeratiivne klasteranalüüs, millega saab vaadelda üheaegselt tuhandeid valke, on muutunud standardseks ja üheks kõige populaarsemaks arvutuslikuks proteoomianalüüsi meetodiks, kuna ühte klastrisse koondunud tundmatute valkude funktsioone on võimalik järeldada samas klastris asuvate tuntud valkude funktsioonidest. Klasterdamise tulemusi võib kujutada järjestikulistel tasanditel ühinemisi ja jagunemisi visualiseeriva dendrogrammi ehk puu-tüüpi diagrammina, mida saab kasutada ka sisendina edasiseks klasterdamiseks. (Luo jt, 2003.) Hierarhilise klasterdamise puuduseks on ajamahukas arvutuslik keerukus (Zhang, Zhang, 2006).

1.3. Kvantiil-kvantiil graafik

Q-Q graafik ehk kvantiil-kvantiil graafik on graafiline tööriist, mis aitab hinnata, kas andmete kogum on usutav. Seda, kas statistilise analüüsi sõltuv muutuja on normaalselt jaotatud, saab kontrollida Q-Q graafikuga. See on küll kõigest visuaalne ning seega subjektiivne hinnang, mis ei anna täieulatuslikku tõendust andmestiku kehtivuse kohta, kuid võimaldab siiski mingil määral kontrollida, kas püstitatud hüpotees on usutav – kui ei, siis tuleb leida, millisest punktist viga lähtub.

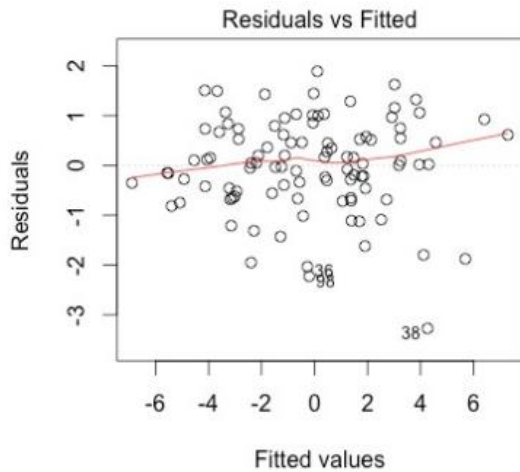
Q-Q graafik on punktgraafik, mis on loodud, kandes saadud andmestiku kaks kvantiili üksteise vastu. Kui mõlemad kvantiilid on ühesuguse jaotusega ehk sarnased, peab joon graafikul olema lineaarne.

Q-Q graafiku eelised on järgmised: proovid ei pea olema võrdsed, paljusid jaotusprobleemide aspekte saab üheaegselt testida, saab leida nihkumise asukohti, nihkeid, muutusi sümmeetrias, saab leida võõrväärtusi, mis andmestikku ei sobi. (<https://data.library.virginia.edu/diagnostic-plots/>)

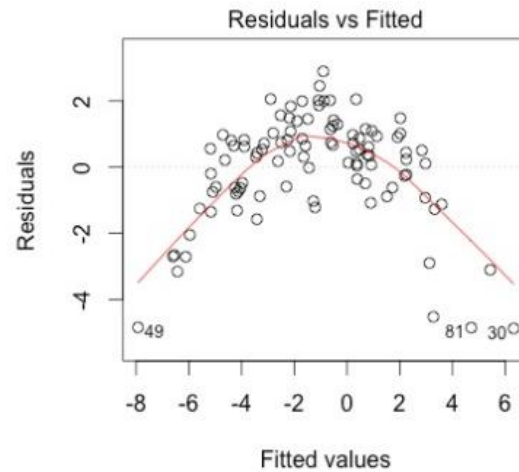
1.4. *Residuals vs fitted*

Residuals vs fitted on graafiline meetod, mis näitab, kas suvalisest andmestikust saadud lineaarne regressioon on tõene. See on hajuvusgraafik, kus y-teljel on jäägid ja x-teljel on *fitted*-väärtused ehk hinnang. Jääkide analüüs mängib olulist rolli prognoosimodeli

kujundamisel. Kõnealune graafik näitab, kas jäägid annavad mittelineaarse mustri. Kui horisontaalsel teljel on võrdsest jaotatud jääke, mis moodustavad selgeid mustreid, siis saab sellest järeldada, et andmestik on mittelineaarne.



Joonis 1.



Joonis 2.

Esimene joonis näitab, et kehtib lineaarne mudel, kus punktid 36, 98 ja 38 kalduvad lineaarsusest kõrvale. Teise joonise peal on näha punast kõverjoont, seega antud mudel ei ole lineaarse seosega.

Fitted-väärtus on teisiti öeldult ennustatud väärtus. On leitud, et *fitted*-väärtuse x saab avaldada lineaarsest regressioonivõrrandist, kus y tähistab ennustatud väärtust:

$$\hat{y} = b_0 + b_1x.$$

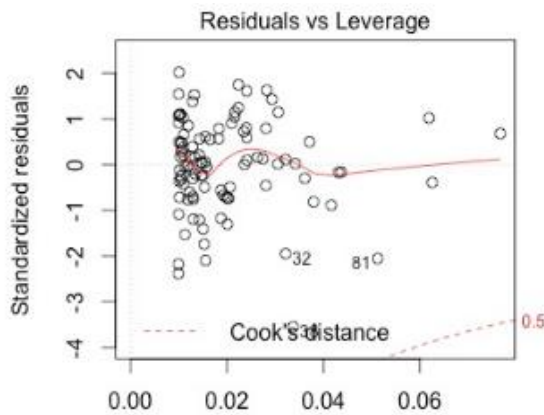
Jääkide arvu tähistatakse e -ga, see on viga ehk kahe vaadeldava objekti vahe ehk ennustatud ehk *fitted*-väärtus. Graafiliselt on see vertikaalne vahemaa andmestikust saadud punkti ja sellesama punkti ennustatava koha väärtuste vahe

$$e = y - \hat{y}$$

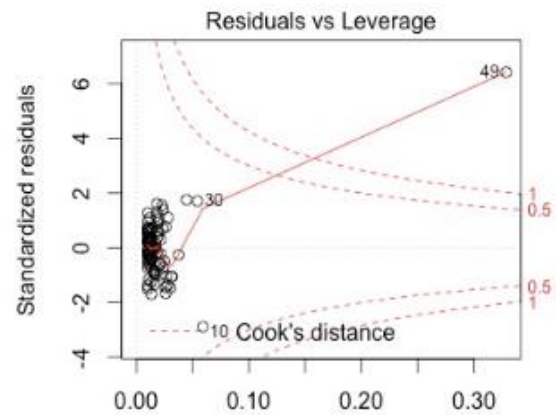
e väärtused on x -teljel ja ennustatud ehk *fitted*-väärtus on y -teljel. (Gipe, 1976.)

1.5. Residuals vs leverage

Leverage on antud kontekstis arv, mis mõõdab regressioonimudeli võimet liikuda iseeneslikult y-telje suunas. *Leverage*'i mõõdetud ennustatud väärtus võib muutuda juhul, kui vaadeldav regressioonimudel nihkub ühe ühiku võrra y-telje suunas. *Leverage*'i arvuline väärtus jääb alati 0 ja 1 vahemikku. Punktil, millel on 0 *leverage*'i, ei ole mõju regressioonimudelile. See graafik aitab leida mõjusid, kui neid on. Kõik võõrväärtused ei mõjuta lineaarset regressiooni. Teistest graafikutest erinevalt ei oma muustrite olemasolu selle graafiku puhul mingit tähtsust – vaadata tuleb äärmisi väärtusi paremal all ja ülal nurgas. Nendel punktidel võib olla suurem mõju regressioonisirgele. Vaadelda tuleb punkte, mis asuvad väljaspool punktiirjoont ehk Cooki kaugust. Väljaspool Cooki kaugust asuvad punktid mõjutavad regressioonitulemusi. Tulemused muutuvad, kui need punktid eemaldada.



Joonis 3.



Joonis 4.

Vasakpoolsel joonisel puuduvad punktid, mis omaksid mõju regressioonisirgele. Parempoolsel joonisel on üks punkt väljaspool Cooki kaugust, see mõjutab regressioonisirget suurel määral. (Gipe, 1976.)

1.6. Korrelatsioonikordaja seos korrigeeritud determinatsioonikordajaga

Korrelatsioonikordaja ehk Pearsoni kordaja, mida märgitakse tähega r , mõõdab kahe muutuva vahelist lineaarset seost. Korrelatsioonikordajat kasutatakse sageli valesti, jättes arvestamata, et sellega ei saa kontrollida, kas mudel on lineaarne. Pearsoni kordaja väärtus jääb vahemikku $+1$ kuni -1 . (Miles, 2014.)

Korrelatsioonikordaja väärtus ruutu võetuna on determinatsioonikordaja. Determinatsioonikordajat kasutatakse, et määrata, kui hästi regressioonivõrrand andmestikku sobib. Korrigeeritud determinatsioonikordaja on tihedalt seotud determinatsioonikordajaga, mida tuntakse kui R -ruutu. R -ruut on statistiline meede, mis näitab, kui suurt osa variatsioonist saab kirjeldada regressioonisirgega. (Samas.)

Korrigeeritud R -ruudu väärtus on alati 0 ja 1 vahemikus. Mida lähemale R -ruut on 1 -le, seda paremini kirjeldab regressioonivõrrand mudelit – X ja Y suhet. Väärtused vahemikus 0 kuni $0,3$ näitavad, et regressioonivõrrand sobib nõrgalt mudelisse, $0,3$ – $0,7$ näitab mõõdukat sobivust ja $0,7$ kuni $1,0$ tugevat. Oluline erinevus R -ruudu ja korrigeeritud R -ruudu vahel on see, et R -ruut suureneb automaatselt, kui lisada andmestikku sõltumatu muutuja, kuid korrigeeritud R -ruudu puhul ei pruugi see juhtuda. Korrigeeritud R -ruut suureneb juhul, kui lisatakse selline sõltumatu muutuja, mis muudab regressioonisirget märgatavalt. Seega on korrigeeritud R -ruut R -ruuduga võrreldes palju kasulikum meede, hindamaks regressioonivõrrandi kattuvust prooviandmestikuga. (Samas.)

1.7. Ekstratsellulaarsete partiklite genereerimine MLV gag valgu abil

Mikrovesiikulid ehk viiruslaadsed partiklid on eksosoomidest suuremad, 100 – 1000 nm läbimõõduga vesiikulid, mida rakud sekreteerivad ekstratsellulaarsesse ruumi. Mikrovesiikulid tekivad rakustressi ajal plasmamembraani väljapoole pungumisel.

Viiruslaadsete partiklite tekkimiseks piisav tingimus on viiruse poolt kodeeritud gag-valgu ekspressioon. Enamiku retroviiruste puhul on gag-molekulid sihitud plasmamembraani, kus need akumulatsioonivad ja organiseeruvad valguliseks võrestikuks, mis edasisel formeerumisel

muutub ebaküpseks viiruslaadseks partiklits ja vabaneb. Seejärel lõhustab viirusproteaaas ebaküpse partikli, muutes selle küpseks. (Datta jt, 2011.)

Mikrovesiikulite tekkimise käigus koguneb nendesse mitmesuguseid raku komponente. Mitmesugustest rakutüüpidest pärinevate struktuuridena erinevad mikrovesiikulid üksteisest molekulaarse koostise poolest. Mikrovesiikuleid liigitatakse selle järgi, milline rakk neid sekreteerib – vereliistakud, punalibled või endoteeli rakud. (Pocsfalvi jt, 2015.)

Mikrovesiikuli moodustumise ajal tekib fosfolipiidide ümberjaotamise ja tsütoskeleti valkude kokkutõmbumise tulemusena dünaamiline vastasmõju. Vesiikulite moodustumine on indutseeritud fosfolipiidide translokatsioonist välimisse membraani. Pungumise protseduur on läbitud, kui tsütoskeleti struktuurid on aktiini ja müosiini vastasmõju toimel kokku tõmbunud. (Samas.)

Uurimuses „Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens” (Kurg jt, 2016) teostati valgusisalduse iseloomustamiseks viiruslaadsetes partiklites LC-MS/MS põhine märgisevaba kvantitatiivne proteoomika analüüs. Antud töös indutseeriti viiruslaadsete partiklite teke MLV gag valgu üleekspressiooniga ja kontrolliti erinevate melanoomi antigeenide paiknemist tekkinud VLP-des. Selle tulemusena tuvastati viiruslaadsetes partiklites kõik viis melanoomi antigeeni, valgud MART1, TRP1, MCAM, MAGEA4 ja MAGEA10. Gag valku oli viiruslaadsetes partiklites kõige rohkem. Lisaks tuvastati kõigis viiruslaadsetes partiklites enam kui viissada unikaalset rakulist valku ning identifitseeriti mitu MLV gag valguga interakteeruvat valku, mis on olulised viiruse pungumisel, nagu Tsg101, Alix, Nedd4 ja klatriini adaptervalgu kompleksid AP2 ja AP1. Enamik LC-MS/MS meetodiga tuvastatud ja rakusiseste valkudega vastavuses olevaid peremeesvalkudest, mis osalevad mitmetes tsütoplasmaatilistes süsteemides nagu tsütoskelett (aktiin, müosiin), mikrotuubulid (tubuliin), tsäperonid, ubikvitiinsõltuv proteosoomisüsteem, valkude ja vesiikulite transport, signalisatsioon ja metabolism, paikneb eeldatavalt viiruslaadsetes partiklites. Tuvastati mitmeid ribosoomivalke, translatsioonifaktoreid, RNA helikaase, hnRNP valke ja tRNA süntetaase. (Kurg jt, 2016.)

2. Eksperimentaalne osa

2.1. Eesmärgid

Käesoleva bakalaaurusetöö eesmärgiks on LFQ intensiivsusi kasutades leida, milliste teiste valkude suhtelist arvukust mikrovesiikulites mõjutab valkude MCAM, MART1, MAGEA10, MAGEA4, MLV gag ja TRP1 üleekspressioon.

Nimetatud eesmärgi saavutamiseks teostati proteoomika andmete statistiline analüüs, kasutades R-studio ja SPSS programmi ning uurimusest „Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens” (Kurg jt, 2016) saadud algandmeid. Eelmainitud üleekspresseeritud valkude iga kombinatsiooni kohta loodi R-studio programmis neli graafikut (regressiooni graafik, Q-Q graafik, *residuals vs fitted* ja *residuals vs leverage*), võrdlemaks viiruslaadsetes partiklites olevate valkude LFQ intensiivsusi. SPSS programmiga loodi aglomeratiivse klasterdamise abil dendrogrammid, näitamaks, millised MCAM, MART1, MAGEA10, MAGEA4, MLV gag ja TRP1 valkude kombinatsioonist mõjutatud valgud on teistega võrreldes LFQ intensiivsuse poolest mikrovesiikulites sarnased.

2.2. Materjal ja meetodika

2.2.1. VLP-de saamine ja mass-spektromeetriline analüüsimine

Melanoomi antigeene kandvate VLP-de tekitamist ja märgisevabal LC-MS/MS baseeruvat kvantitatiivset proteoomianalüüsi on kirjeldatud uurimuses „Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens” (Kurg jt, 2016).

Mass-spektromeetrilisel analüüsil saadud algandmeid vaadeldi, kasutades programmi MaxQuant 1.4.0.8 (samas). Selle tulemusena saadud andmeid valkude suhtelise arvukuse kohta kasutatakse käesolevas bakalaureusetöös, leidmaks valkude MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM seoseid teiste valkude suhtelise arvukusega.

2.2.2. Regressioonanalüüs R-studio programmiga

Kasutasin R-studio programmi valkude MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM seostatuse leidmiseks teiste valkude suhtelise arvukusega mikrovesiikulites LFQ intensiivsuse alusel. R-Studio on arvutiteadlase J. J. Allaire loodud vabavara andmete statistiliseks töötlemiseks ja tööks graafikutega.

Melanoomi kõigi nimetatud antigeenidega tehti esmalt kolm katset viiruslaadsetes partiklites olevate valkude LFQ intensiivsuse leidmiseks ja seejärel arvutati programmis Excel nende kolme katse keskmine, kasutades käsklust =AVERAGE. Niiviisi saadud andmestik on salvestati arvutisse CSV formaadis (*Comma-separated values*) ja seejärel kanti plottimiseks uuesti üle R-studio programmi. Eelmainitud üleekspressioneeritud valkude iga kombinatsiooni kohta loodi R-studio programmis neli graafikut (regressiooni graafik, Q-Q graafik, *residuals vs fitted* ja *residuals vs leverage*), võrdlemaks viiruslaadsetes partiklites olevate valkude LFQ intensiivsusi. Edasi valiti saadud graafikute hulgast välja need, mille R-ruut väärtus oli üle 70% ja p-väärtus alla 0,05 ning mille valkude LFQ intensiivsused olid seega vastastikku lineaarses sõltuvuses.

2.2.3. Hierarhiline klasterdamine proteoomianalüüsi mustrite leidmiseks

SPSS ehk Statistical Package for the Social Sciences on tarkvarapakett, mida kasutatakse statistiliseks analüüsiks. Esmalt kanti SPSS programmi iga valgu keskmine LFQ intensiivsus. Edasise alt üles ehk aglomeratiivse klasterdamise käigus kasutati keskmise seostatuse ehk kaalumata keskmise seostatusega meetodit, kuna tegemist oli rohkem kui kahe objekti seosega ning keskmise seostatuse meetod peaks andma klastrite vahelise kauguse kohta kõige täpsema hinnangu. Klasterdamise tulemusena saadi dendrogrammid proteoomianalüüsi mustrite leidmiseks.

2.3. Tulemused ja arutelu

2.3.1. Proteoomide võrdlus, kasutades Q-Q graafikute meetodit

Q-Q graafikud võrdlevad konkreetsete kombinatsioonide andmestike hajuvust omavahel, tuvastades, kas jäägid ehk prognoosivead on normaaljaotuse ehk teoreetiliselt eeldatud jaotusega. Numbritega märgistatud punktid tähistavad valke, mille LFQ intensiivsus väga suure tõenäosusega ei lange normaaljaotusega kokku. Tabel 1 näitab, milliste kombinatsioonide andmestike millised punktid (valgud) ei ole kahe andmestiku omavahelise teoreetilise jaotuse hajuvusega kooskõlas. Q-Q graafikud kõikide kombinatsioonide kohta on toodud lisas (joonised 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76).

Kõnealuse tabeli alusel saame väita, et kombinatsioonides MAGEA10xMAGEA4, MAGEA10xMART1, MAGEA10xMCAM, MAGEA10xMLV gag ja MAGEA10xTRP1 ei ole teoreetilise jaotuse hajuvusega kooskõlas punktid 336 ja 372 (vastavalt 'Histone H3' ja 'Histone H2A type 1-H'), kombinatsioonides MAGEA10xMART1, MAGEA10xMCAM, MAGEA4xMCAM ja MAGEA4xTRP1 punkt 181 ('Histone H2B type 1-P'), kombinatsioonides MAGEA10xMAGEA4 punkt 162 ('Serine/threonine-protein phosphatase PP1-alpha catalytic subunit'), kombinatsioonides MAGEA10xTRP1 punkt 182 ('Histone H2B type 1-C/E/G'), kombinatsioonides MAGEA4xMART1, MAGEA4xMCAM ja MART1xMCAM punkt 656 ('Alpha-2-macroglobulin-P'), kombinatsioonides MAGEA4xMCAM, MAGEA4xMLV gag, MART1xMCAM ja MCAMxTRP1 punkt 10 ('Heat shock cognate 71 kDa protein'), kombinatsioonides MART1xMCAM ja MLV gag xTRP1 punkt 20 ('GAG MLVFF'), kombinatsioonides MART1xMLV gag, MCAMxMLV gag ja MCAMxTRP1 punkt 34 ('Tubulin beta-4B chain'), kombinatsioonides MART1xMLV gag ja MCAMxMLV gag punkt 12 ('Pyruvate kinase isozymes M1/M2'), kombinatsioonides MART1xTRP1 punktid 50 ja 60 (vastavalt 'Actin, cytoplasmic 1' ja 'Glyceraldehyde-3-phosphate dehydrogenase'), kombinatsioonides MLV gag xTRP1 punkt 484 ('Osteopontin').

Tabel 1. Teoreetilisest normaaljaotusest kõrvale kalduva LFQ intensiivsusega valkudele vastavate geenide nimetused.

	MAGEA10	MAGEA4	MART1	MCAM	MLV gag	TRP1
MAGEA10	x					
	Hist1h4a, Hist1h3b, Hist1h2ah					
MAGEA4	162, 336, 372	x				
	Hist1h2bc, Hist1h3b, Hist1h2ah	Hist1h2bc, A2mp				
MART1	181, 336, 372	181, 656	x			
	Hist1h2bc, Hist1h3b, Hist1h2ah	Hspa8, Hist1h2bc, A2mp	Hspa8, GAG MLVFF, A2mp			
MCAM	181, 336, 372	10, 181, 656	10, 20, 656	x		
	Hist1h3b, Hist1h2ah	Hspa8, Pkm	Pkm, Tubb4b	Pkm, Tubb4b		
MLV gag	336, 372	10, 12	12, 34	12, 34	x	
	Hist1h2bc, Hist1h3b, Hist1h2ah	Hist1h2bc, A2mp	Actb, Gapdh	Hspa8, Tubb4b	GAG MLVFF, Spp1	
TRP1	181, 336, 372	181, 656	50, 60, 656	10, 34	20, 484	x

2.3.2. Regressioonigraafikute analüüs

Regressioonigraafik näitab kombinatsiooni andmetike sobivust regressioonanalüüsi mudelisse. Mida suurem on korrigeeritud R-ruudu väärtus, seda suurem on võrreldavate andmetike omavaheline korrelatsioon.

Tabeli 2 alusel saab öelda, et kombinatsioonide MAGEA10xMAGEA4, MAGEA10xMART1, MAGEA10xMCAM, MAGEA10xTRP1, MAGEA4xMART1, MAGEA4xMCAM, MAGEA4xTRP1, MART1xMCAM, MART1xTRP1 ja MCAMxTRP1 andmetike omavaheline korrelatsioon on piisav, määratlemaks lineaarset sõltuvust.

Tabel 2. Regressioonigraafikute korrigeeritud R-ruut väärtused.

	MAGEA10	MAGEA4	MART1	MCAM	MLV gag	TRP1
MAGEA10	x					
MAGEA4	0.8678	x				
MART1	0.8188	0.9443	x			
MCAM	0.7181	0.8945	0.955	x		
MLV gag	0.362	0.3226	0.4403	0.2971	x	
TRP1	0.7591	0.9016	0.9766	0.9794	0.3869	x

2.3.3. *Residuals vs fitted* graafikute analüüs

Residuals vs fitted graafikud näitavad, milliste valkude LFQ intensiivsus erineb ennustatavast LFQ intensiivsusest kõige enam. Niisugune kõrvalekalle näitab, et rakkudesse viidud plasmiididest kodeeritud valgud on väga suure tõenäosusega mõjutanud ennustatavast LFQ väärtusest erineva LFQ intensiivsusega valkude suhtelist arvukust. Graafikutel on konkreetse vaadeldava korrelatsiooni mudeli teoreetiliselt eeldatav joongraafik märgitud punase joonega, numbritega märgistatud punktid tähistavad valke, mille LFQ intensiivsus ennustatud regressioonimudelil olulisel määral erineb. *Residuals vs fitted* graafikud kõikide kombinatsioonide kohta on toodud lisa (joonised 7, 12, 17, 22, 27, 32, 37, 42, 47, 52, 57, 62, 67, 72, 77).

Tabeli 3 alusel saame väita, et kombinatsioonides MAGEA10xMAGEA4, MAGEA10xMART1, MAGEA10xMCAM, MAGEA10xMLV gag ja MAGEA10xTRP1 erineb ennustatavast LFQ intensiivsusest kõige enam, olles ennustatust märgatavalt suurem, punkt 336 ('Histone H3'), kombinatsioonides MAGEA10xMAGEA4, MAGEA10xMART1, MAGEA10xMCAM, MAGEA10xMLV gag, MAGEA10xTRP1 ja MART1xMCAM punkt 372 ('Histone H2A type 1-H'), kombinatsioonides MAGEA10xMART1, MAGEA10xMCAM, MAGEA4xMCAM, MAGEA10xTRP1 ja MAGEA4xTRP1 punkt 181 ('Histone H2B type 1-P'), kombinatsioonis MAGEA10xMAGEA4 punkt 162 ('Serine/threonine-protein phosphatase PP1-alpha catalytic subunit'), kombinatsioonides MAGEA4xMART1, MAGEA4xMCAM, MAGEA4xTRP1, MART1xMCAM, MART1xTRP1 ja MLV gag xTRP1 punkt 656 ('Alpha-2-macroglobulin-P'), kombinatsioonides MAGEA4xMCAM, MAGEA4xMLV gag ja MCAMxTRP1 punkt 10 ('Heat shock cognate 71 kDa protein'), kombinatsioonides MAGEA10xMLV gag, MAGEA4xMLV gag, MAGEA4xTRP1, MART1xMCAM, MART1xMLV gag, MCAMxMLV gag, MCAMxTRP1 ja MLV gag xTRP1 punkt 20 ('GAG MLVFF'), kombinatsioonides MART1xMLV gag ja MCAMxMLV gag punkt 34 ('Tubulin beta-4B chain'), kombinatsioonis MART1xTRP1 punktid 50 ja 60 (vastavalt 'Actin, cytoplasmic 1' ja 'Glyceraldehyde-3-phosphate dehydrogenase'), kombinatsioonis MLVgag xTRP1 punkt 484 ('Osteopontin'). Kombinatsioonides MART1xMLV gag ja

MCAMxMLV gag erineb ennustatavast LFQ intensiivsusest kõige enam, olles ennustatust märgatavalt väiksem, punkt 12 ('Pyruvate kinase isozymes M1/M2').

Tabel 3. Ennustatud regressioonimudelidest oluliselt erineva LFQ intensiivsusega valkudele vastavate geenide nimetused.

	MAGEA10	MAGEA4	MART1	MCAM	MLV gag	TRP1
MAGEA10	X Histone H4, Hist1h3b Hist1h2ah					
MAGEA4	162, 336, 372 Hist1h2bc, Hist1h3b Hist1h2ah	X Hist1h2bc, A2mp				
MART1	181, 336, 372 Hist1h2bc, Hist1h3b Hist1h2ah	181, 656 Hspa8, Hist1h2bc, A2mp	X GAG MLVFF, Hist1h2ah, A2mp			
MCAM	181, 336, 372 GAG MLVFF, Hist1h3b Hist1h2ah	10, 181, 656 Hspa8, Pkm, GAG MLVFF	20, 372, 656 Pkm, GAG MLVFF, Tubb4b	X Pkm, GAG MLVFF, Tubb4b		
MLV gag	20, 336, 372 Hist1h2bc, Hist1h3b, Hist1h2ah	10, 12, 20 GAG MLVFF, Hist1h2bc, A2mp	12, 20, 34 Actb, Gapdh, A2mp	12, 20, 34 Hspa8, GAG MLVFF, Tubb4b	X GAG MLVFF, Spp1 A2mp	
TRP1	181, 336, 372	20, 181, 656	50, 60, 656	10, 20, 34	20, 484, 656	X

2.3.4. Residuals vs leverage graafikute analüüs

Residuals vs leverage graafikud võimaldavad tuvastada erandeid (näiteks sisestusvigu). X-teljel on vaadeldavate punktide mõjukus (*leverage*) – mida suurem on mõjukus, seda enam suudab antud punkt regressioonikõverat enda juurde "painutada". Y-teljel on jääk ehk vaadeldava punkti (LFQ intensiivsuse) erinevus valgu teoreetiliselt eeldatavast LFQ intensiivsusest. Need graafikud näitavad, milliste konkreetsete punktide eemaldamisel muutuks valkude LFQ intensiivsuse regressioonimudel olulisel määral. Kui *residuals vs fitted* graafikud näitavad, millised punktid (valgud) erinevad teoreetiliselt ennustatud regressioonimudelidest kõige rohkem, siis *residuals vs leverage* graafikud aitavad leida valke, mille LFQ intensiivsus mõjutab ennustatud regressioonimudelit kõige tugevamalt.

Residuals vs leverage graafikutes tuleb vaadelda punkte, mis asuvad väljaspool punktiirjoont ehk Cooki kaugust. Cooki kaugus võimaldabki identifitseerida võõrväärtusi ehk väärtusi, mis ei sobi antud regressioonianalüüsi ning esialgset regressioonimudelit oluliselt mõjutavad, seega saab järeldada, et kõige enam on mikrovesiikulisse lisatud valgud mõjutanud nende valkude LFQ intensiivsust, mis jäävad väljapoole Cooki kaugust. *Residuals vs leverage* graafikud kõikide kombinatsioonide kohta on toodud lisas (joonised 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58, 63, 68, 73, 78).

Tabel 4. Ennustatud regressioonimudelit kõige enam mõjutava LFQ intensiivsusega valkudele vastavate geenide nimetused.

	MAGEA10	MAGEA4	MART1	MCAM	MLV gag	TRP1
MAGEA10	x Actb, Hist1h2ah, Hist1h3b					
MAGEA4	50, 372, 336 GAG MLVFF, Hist1h3b, Hist1h2ah	x Hspa8, GAG MLVFF, Actb				
MART1	20, 336, 372 Hspa8, Eef1a1, Hist1h2ah	10, 20, 50 Hspa8, Eef1a1, Actb	Hspa8, GAG MLVFF, Actb			
MCAM	10, 49, 372 Hspa8, GAG MLVFF, Tuba1c	10, 49, 50 Hspa8, GAG MLVFF, Actb	10, 20, 50 Hspa8, GAG MLVFF, Actb	x		
MLV gag	10, 20, 25 Hspa8, GAG MLVFF, Hist1h2ah	10, 20, 50 Hspa8, GAG MLVFF, Actb	10, 20, 50 Pkm, Actb, Gapdh	10, 20, 25 Hspa8, GAG MLVFF, Tuba4b	x Pkm, GAG MLVFF, Gapdh	
TRP1	10, 20, 372	10, 20, 50	12, 50, 60	10, 20, 34	12, 20, 50	x

Tabeli 4 alusel saame väita, et valk MAGEA10 puhul jääb peaaegu igas antud kombinatsioonis väljapoole Cooki kaugust punkt 372 ('Histone H2A type 1-H'), välja arvatud kombinatsioonis MAGEA10xMLV gag; MAGEA4 puhul jääb iga antud kombinatsiooni tulemusena Cooki kaugusest väljapoole punkt 50 ('Actin, cytoplasmic 1') ja kõigis kombinatsioonides peale kombinatsiooni valguga MAGEA10 ka punkt 20 ('GAG MLVFF'); MCAM puhul jääb iga antud kombinatsiooni tulemusena Cooki kaugusest

väljapoole punkt 10 ('Heat shock cognate 71 kDa protein'), kombinatsioonides valkudega MART1, MLV gag ja TRP1 ka punkt 20 ('GAG MLVFF') ning kombinatsioonides valkudega MAGEA10 ja MAGEA4 ka punkt 49 ('Elongation factor 1 alpha 1'); MLV gag puhul jääb igas antud kombinatsioonis väljapoole Cooki kaugust punkt 20 ('GAG MLVFF'), kombinatsioonis kõigi valkudega peale TRP1 ka 10 ('Heat shock cognate 71 kDa protein'), kombinatsioonis valkudega MAGEA4, MART1 ja MLV gag punkt 50 ('Actin, cytoplasmic 1') ning valkudega MAGEA10 ja MCAM punkt 25 ('Tubulin alpha-1C chain').

2.3.5. Hierarhilise klasterdamise tulemused

MAGEA4 ja MAGEA10 LFQ intensiivsuste hierarhilise klasterdamise tulemuseks on dendrogramm, mis näitab, millised nende kombinatsioonist mõjutatud valgud on suurema tõenäosusega LFQ intensiivsuse poolest sarnased. Valkudest, mille puhul regressioonanalüüs näitas, et nende LFQ intensiivsust mõjutavad valgud MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM, sarnanevad dendrogrammi kohaselt LFQ intensiivsuse poolest omavahel 'Heat shock cognate 71 kDa protein', 'Hemoglobin subunit beta-2', 'Actin, cytoplasmic 1' ja 'Histone H1.1'. Nendega sarnaneb ka valk 'Alpha-enolase', mille mõjutatust valkudest MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM regressioonanalüüs ei näidanud. (Vt lisa 2, joonis 80.)

MAGEA4 ja MART1 LFQ intensiivsuste hierarhilise klasterdamise tulemusena saadud dendrogrammist on näha, et nende kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel valgud 'Heat shock cognate 71 kDa protein', 'Elongation factor 1-alpha 1', 'Histone H2A type 1-H', 'Hemoglobin subunit beta-2' ja 'Serotransferrin'. (Vt lisa 2, joonis 81.)

MAGEA4 ja MCAM kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel valgud 'Actin, cytoplasmic 1', 'Heat shock protein 105 kDa', 'T-complex protein 1 subunit alpha' ja 'Histone H2A type 1-H'. (Vt lisa 2, joonis 82.)

MAGEA10 ja MART1 kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel valgud 'Glyceraldehyde-3-phosphate dehydrogenase', 'Tubulin beta-4B chain' ja 'Elongation factor 1-alpha 1'. (Vt lisa 2, joonis 83.)

MAGEA10 ja MCAM kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel 'Hemoglobin subunit beta-2', 'Glyceraldehyde-3-phosphate dehydrogenase', 'Tubulin beta-4B chain' ja 'Serotransferrin'. (Vt lisa 2, joonis 84.)

MAGEA10 ja TRP1 kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel 'Heat shock cognate 71 kDa protein', 'Elongation factor 1-alpha 1' ja 'Histone H2A type 1-H'. (Vt lisa 2, joonis 85.)

MART1 ja TRP1 kombinatsioonist mõjutatud valkudest sarnanevad LFQ intensiivsuse poolest omavahel valgud 'Heat shock cognate 71 kDa protein', 'Histone H4' ja 'Hemoglobin subunit beta-2'. (Vt lisa 2, joonis 86.)

Tabel 5. Geenid, mis on vastavuses rakkudesse viidud plasmiididest kodeeritud niisuguste valkudega, mis on melanoomi antigeenide kombinatsioonide puhul LFQ intensiivsuse poolest omavahel kõige enam sarnased.

	MAGEA10	MAGEA4	MCAM	MART1	MLV GAG	TRP1
MAGEA10	x					
MAGEA4	Hspa8,Hbb-b2,Eno1,Actb,Hist1h2ah,Hist1h3b,Tubb4b,Eef1a1	x				
MCAM	Hbb-b2,Gapdh,Actb,Eef1a1,Hsp90ab1,Tubb4b,Tubalc	Hspa8,Actb,Hsp90ab1,Eef1a1	x			
MART1	Gapdh,Hsp90ab1,Tubb4b,Eef1a1,Actb	Hspa8,Ptkm,Eef1a1,GAG MLVFF,Eno1	Hspa8,Tubb4b,Eno1,Cct4,Eef2	x		
MLV GAG	GAG MLVFF,Spp1	GAG MLVFF,Gapdh,Tubb4b	GAG MLVFF,Ptkm,Tubb4b,A2mp	GAG MLVFF,Ptkm,Tubb4b,A2mp	x	
TRP1	Hspa8,Hist1h2ah,Hsp90ab1,Eef2,Eef1a1,Hbb-b2,Gapdh,Tubb4b	GAG MLVFF,Actb,Tubalc,Eno1	Actb,Eef2,Tubb4b,Hbb-b2,GAG MLVFF,Eef2,Tubalc	Hspa8,Gapdh,Eno1,GAG MLVFF,Ptkm	GAG MLVFF,Ptkm,Tubb4b,Uba52,Hbb-b2	x

2.3.6. Arutelu

Nii Q-Q kui ka *residuals vs leverage* ja *residuals vs fitted* graafikute alusel saame väita, et MAGEA10 üleekspressiooni puhul on punktid 336 ehk 'Histone H3' ja 372 ehk 'Histone H2A type 1-H' MAGEA10 partiklites enam esindatud kui teisi melanoomi antigeene sisaldavates VLP-des. H3 ja H2A on nukleosoomi ühed olulisemad komponendid, mis muudavad DNA kompaktsemaks ja piiravad teatud valkude ligipääsu, mis vajavad DNA-d matriitsina. (Tagami jt, 2004.) Seega võib eeldada, et MAGEA10 partiklid võivad sisaldada nukleosoomi.

Punkt 181 ehk 'Histone H2B type 1-P' puhul näitavad nii Q-Q kui ka *residuals vs fitted* graafikud, et selle valgu esindatus VLP-des on nii MAGEA10 kui MAGEA4 üleekspressiooni korral oluliselt kõrgem. Seega võib eeldada, et MAGEA10 ja MAGEA4 üleekspressiooni toimel hakkab rakk juurde tootma tsütosoolse sensorina toimivat ekstrakromosomaalset histooni H2B, mille abil tuvastatakse kaheaheelalisi DNA fragmente. Niisugused fragmendid võivad kuuluda infektsioonitekitajatele rakus või olla tekkinud rakulise kahjustuse tõttu. Võõr-DNA sissetungimise või rakukahjustuse tulemusena aktiveerib H2B interferoon (IFN)- β promootori, millega kaasneb immuunvastus. (Kobiyama, K. Jt, 2013.) Histooni H2B suhtelise arvukuse suurenemine VLP-des MAGEA10 ja MAGEA4 toimel võib tähendada, et melanoomi antigeenid MAGEA10 ja MAGEA4 võivad rakku kahjustada.

Nagu Q-Q ja *residuals vs fitted* graafikutest näha, tõstab nii MAGEA4 kui MART1 üleekspressioon suure tõenäosusega ka proteaasi inhibiitorina tuntud ja teatavate kasvufaktoritega seonduva 'Alpha-2-macroglobulin-P' (punkt 656) suhtelist arvukust VLP-des. Ekstratsellulaarne valk 'Alpha-2 macroglobulin' (A2M) säilitab rakus kasvufaktorite ja tsütokineesi homöostaasi, inaktiveerib proteaase ning metalloproteaase, mis lagundavad kasvajate metastaseerumisel rakumembraani. (Smorenburg jt, 1996.) Kuna 'Alpha-2 macroglobulin' on peamine proteaasi inhibiitor, siis võib selle suhtelise arvukuse suurenemine VLP-des MAGEA4 ja MART1 toimel tähendada, et need melanoomi antigeenid võivad suurendada proteaaside ja metalloproteaaside kogust.

Valkude MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM LFQ intensiivsuse andmestiku alusel *residuals vs leverage* graafikute Cooki kaugusega saadud tulemustest võib järeldada, et kõik need valgud tõstavad punkt 10 ehk valk 'Heat shock cognate 71 kDa protein' (Hsp70) suhtelise arvukuse taset VLP-des. Vähesel määral kinnitavad sellist seost ka Q-Q ja *residuals vs fitted* graafikud. Hsp70 kuulub kuumašokivalkude hulka, mis toimivad tšaperonidena raku teiste valkude voltimisel ja korrastamisel ning valesti voltunud valkude uuesti voltimisel. Hsp70 valkude suhteline arvukus suureneb stressiolukorras. (Mayer jt, 2005). Saadud tulemuste põhjal võib järeldada, et kõik eespool loetletud valgud võivad põhjustada rakustressi.

Residuals vs leverage graafikute tulemustest võib järeldada, et melanoomi antigeenid MAGEA10, MAGEA4, MART1, TRP1 ja MCAM tõstavad kõik punkti 50 ehk valgu 'Actin, cytoplasmic 1' (ACTB) suhtelist arvukust VLP-des. Mitmed uuringud on näidanud, et ACTB polümeerisatsioon ja lokaliseerumine võivad soodustada vähirakkude liikuvust, invasiivsust ja metastaseerumist. ACTB polümeerisatsiooni ja tsütoskeleti moodustumise käigus tekivad vähirakul väljasopistused, mis aitavad rakul liikuda. (Guo jt, 2012). Kuna ACTB üleekspressioon suurendab vähirakkude liikuvust, siis võib sellest järeldada, et nii MAGEA10, MAGEA4, MART1, TRP1 kui ka MCAM soodustavad kasvajate metastaseerumist.

Q-Q ja *residuals vs fitted* graafikute alusel saab väita, et TRP1 üleekspressiooni puhul on punkt 484 ehk 'Osteopontin' TRP1 partiklites enam esindatud kui teisi melanoomi antigeene sisaldavates VLP-des. 'Osteopontin' mängib olulist rolli vähi progresseerumise, metastaseerumise ja angiogeneesi juures, indutseerides onkogeene, suurendades metastaatilist potentsiaali mitmesugustes vähirakkudes ja soodustades nende rakkude ellujäämist (Ahmed jt, 2011). Eelnevast võib järeldada, et kuna TRP1 üleekspressioon võib põhjustada 'Osteopontini' suhtelise arvukuse tõusu VLP-des, siis võib see omakorda soodustada vähirakkude ellujäämist ja metastaseerumist.

Nii Q-Q kui ka *residuals vs leverage* ja *residuals vs fitted* graafikute alusel on näha, et TRP1 üleekspressioon tõstab punkt 60 ehk valgu 'Glyceraldehyde-3-phosphate

dehydrogenase' (GAPDH) suhtelist arvukust. GAPDH on multifunktsionaalne valk, mis osaleb glükolüüsil ja apoptoosis ning võib glükolüütilist ainevahetust suurendades põhjustada kasvaja progresseerumist ja metastaatilist levikut (Ramos jt, 2015). Kuna TRP1 üleekspressioon võib põhjustada GAPDH suhtelise arvukuse tõusu VLP-des, siis võib see omakorda soodustada vähirakkude metastaatilist levikut.

Punkt 20 on MLV gag valk ise, mis on kõigis VLP-des alati ja igasugustel tingimustel üks suhteliselt suurima arvukusega valkudest (Kurg jt, 2016).

Lineaarset sõltuvust LFQ intensiivsuste vahel näitas regressioonanalüüs. Peaaegu täielik lineaarne sõltuvus (R-ruut väärtus üle 0,7) on peaaegu kõigi kombinatsioonide vahel, välja arvatud kombinatsioonid valguga MLV gag (R-ruut väärtused alla 0,5).

Statistilist olulisust väljendav 0 ja 1 vahele jääv p-väärtus, mille piiriks loetakse tulemust $p=0,05$, oli kõigi kombinatsioonide puhul väiksem kui $2,2 \times 10^{-16}$.

Regressioonanalüüsiga leitud valgud, mille LFQ intensiivsust on mõjutanud valgud MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM, on osaliselt näha ka hierarhilise klasterdamisega loodud dendrogrammides. Nende valkudega on omavahel sarnase LFQ intensiivsuse alusel seotud ka niisuguseid valke, mida regressioonanalüüs ei leidnud. Nimetatud seos võib tähendada, et kõik leitud valgud võivad kuuluda ühte kompleksesse süsteemi, mida tuleks edaspidi uurida.

Kokkuvõte

Käesoleva bakalaaurusetöö eesmärk on leida melanoomi antigeenidega MAGEA10, MAGEA4, MART1, TRP1, MLV gag ja MCAM rikastatud VLP-de koostise sõltuvus nimetatud valkudest ja seosed teiste valkude suhtelise arvukusega, kasutades LFQ intensiivsust R-studio ja SPSS programmis.

R-studio programmis läbi viidud regressioonanalüüsi tulemustest saab järeldada, et valgu 'Heat shock cognate 71 kDa protein' LFQ intensiivsust tõstavad valgud MAGEA10, MAGEA4, MLV gag, TRP1, MART1 ja MCAM, mis võivad seega kõik põhjustada rakustressi, kuna Hsp70 valkude suhteline arvukus suureneb stressiolukorras.

MAGEA10 üleekspressiooni puhul on valgud 'Histone H3' ja 'Histone H2A type 1-H' MAGEA10 partiklites enam esindatud kui teisi melanoomi antigeene sisaldavates VLP-des. Kuna H3 ja H2A on nukleosoomi ühed olulisemad komponendid, siis võib eeldada, et MAGEA10 partiklid võivad sisaldada nukleosoomi.

'Histone H2B type 1-P' esindatus VLP-des on nii MAGEA10 kui MAGEA4 üleekspressiooni korral oluliselt kõrgem. Seega võib eeldada, et MAGEA10 ja MAGEA4 üleekspressiooni toimel hakkab rakk juurde tootma tsütosoolse sensorina toimivat ekstrakromosomaalset histooni H2B, mille suhtelise arvukuse suurenemine VLP-des võib tähendada, et melanoomi antigeenid MAGEA10 ja MAGEA4 võivad rakku kahjustada.

Nii MAGEA4 kui MART1 üleekspressioon tõstab suure tõenäosusega proteaasi inhibiitorina tuntud 'Alpha-2-macroglobulin-P' suhtelist arvukust VLP-des, mis võib tähendada, et kõnealused melanoomi antigeenid võivad suurendada proteaaside ja metalloproteaaside kogust. Proteaaside ja metalloproteaaside üks funktsioone on rakumembraani lagundamine kasvajate metastaseerumisel.

Melanoomi antigeenid MAGEA10, MAGEA4, MART1, TRP1 ja MCAM tõstavad kõik valgu 'Actin, cytoplasmic 1' (ACTB) suhtelist arvukust VLP-des. Kuna ACTB polümerisatsioon ja lokaliseerumine võivad soodustada vähirakkude liikuvust ja

invasiivsust, siis võib sellest järeldada, et nii MAGEA10, MAGEA4, MART1, TRP1 kui ka MCAM võivad soodustada kasvajate metastaseerumist.

TRP1 üleekspressioon tõstab suure tõenäosusega nii valgu 'Osteopontin' kui valgu 'Glyceraldehyde-3-phosphate dehydrogenase' (GAPDH) suhtelist arvukust VLP-des. Kuna 'Osteopontin' mängib olulist rolli vähi progresseerumise, metastaseerumise ja angiogeneesi juures, GAPDH aga võib glükolüütilist ainevahetust suurendades põhjustada kasvajate progresseerumist ja metastaatilist levikut, siis võib eeldada, et TRP1 võib soodustada vähi progresseerumist ja metastaatilist levikut.

SPSS programmis teostatud aglomeratiivse klasteranalüüsi eesmärk oli liigitada objektid organiseeritud süsteemi, kus igas rühmas on sarnaste omadustega liikmed. Regressioonanalüüsiga leitud valgud on osaliselt näha ka hierarhilise klasterdamisega loodud dendrogrammides, kus nendega on omavahel sarnase LFQ intensiivsuse alusel seotud ka niisuguseid valke, mida regressioonanalüüs ei leidnud.

Nimetatud seos võib tähendada, et kõik valgud, mille puhul seosed leiti, võivad kuuluda ühtsesse kompleksesse süsteemi, mida tuleks edaspidi uurida.

Proteomic analysis of virus-like particles carrying melanoma antigens

Ranel Tammela

Summary

Previous works of our lab had shown that different melanoma antigens, MAGEA10, MAGEA4, MART1, TRP1, MLV gag and MCAM are incorporated into virus-like particles induced by over-expression of MLV Gag protein. The aim of this bachelor thesis is to find the links between these melanoma antigens with other protein abundances by analyzing their LFQ intensities with R-studio and SPSS program.

R-studio carried out regression analysis and it can be concluded that the protein, 'Heat shock cognate 71 kDA', abundance can be regulated by MAGEA10, MAGEA4, MLV gag, TRP1, MART1, and MCAM. The reason for this multitude might be due to the fact that previously mentioned six proteins might cause stress in cells, because Hsp70 proteins relative abundance increases with stress, respectively.

Overexpression of MAGEA10 causes 'Histon H3' and 'Histone H2A type 1-h' to be presented with higher abundancy in virus-like particles. Since H3 and H2A are part one of the most important component of a nucleosome. It can be assumed that particles with MAGEA10 proteins contain nucleosomes.

'Histone H2B type1 1-p' is a represented in virus-like particles due to overexpression of MAGEA10 and MAGEA4 and the abundance was substantially higher. Thus, it can be assumed that the overexpression of MAGEA4 and MAGEA10 causes cells to produce a sensor like functioning extrachromosomal cytosolic histone H2B, which may be caused by the damage that previously mentioned MAGE antigens did.

MAGEA4 and MART1 overexpression increases the likelihood of the relative abundance of a protease inhibitor known as 'Alpha-2-macroglobulin-P', which may mean that the mentioned melanoma antigens increase with great probability the quantities of proteases

and metalloproteases. One function regarding proteases and metalloproteases is that those disintegrate cell membranes with the process of metastasis.

All five melanoma antigens (MAGEA4, MART1, TRP1, and MCAM) raise the relative abundance of 'Actin,cytoplasmic 1' (ACTB) in virus-like particles. Since ACTB polymerization and localization may promote cancer mobility and invasiveness. Thus, it can be concluded that MAGEA10, MAGEA4, MART1, TRP1 and MCAM may promote tumor metastasis.

TRP1 overexpression likely increases the abundance of 'Osteopontin' and 'Glyceraldehyde-3-phosphate dehydrogenase' (GAPDH) in virus-like particles. 'Osteopontin' plays an important role in tumor progression, metastasis, and angiogenesis, and GAPDH, however, can lead to tumor progression by increasing the glycolytic metabolism and metastatic spread. Consequently, it can be assumed that TRP1 may promote the progression of cancer and metastatic spread.

SPSS program conducted an agglomerative cluster analysis which was designed to classify an object in an organized system, where members of each group have similar characteristics. The proteins found in the regression analysis were also seen in the dendrograms which was made by hierarchical clustering. These links may mean that all of the proteins from which the links were found, could be included in a single system in a complex manner that needs further investigation.

Tänuõnad

Täna *PhD* Reet Kurge ja *PhD* Aare Abroid juhendamise eest.

Kasutatud kirjandus

- Ahmed, M., Behera, R., Chakraborty, G., Jain, S., Kumar, V., Sharma, P., Bulbule, A., Kale, S., Kumar, S., Mishra, R., Raja, R., Saraswati, S., Kaur, R., Soundararajan, G., Kumar, D., Thorat, D., Sanyal, M., Ramdasi, A., Pompom, G., Kundu, G. C. (2011). Osteopontin: a potentially important therapeutic target in cancer. *Expert Opin. Ther. Targets*, 15(9): 1113-1126.
- Aviner, R., Geiger, T., Elroy-Stein, O. (2014). Genome-wide identification and quantification of protein synthesis in cultured cells and whole tissues by puromycin-associated nascent chain proteomics (PUNCH-P). *Nature Protocols*, Vol. 9, 4, 751-760.
- Brazma, A., Vilo, J. (2000). Gene expression data analysis. *FEBS Letters* 480, 17-24.
- Cheon, S. A., Han, E. J., Kang, H. A., Ogrydziak, D. M., Kim, J.-Y. (2003) Isolation and characterization of the TRP1 gene from the yeast *Yarrowia lipolytica* and multiple gene disruption using a TRP blaster. *Yeast*, Vol. 20, 8, 677-685.
- Cichosz, P. 2015. *Data Mining Algorithms*, p. 349-372. John Wiley & Sons, Ltd.
- Clough, T., Thaminy, S., Ragg, S., Aebersold, R., Vitek, O. (2012). Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. *BMC Bioinformatics*, 13(Suppl 16):S6.
- Datta, S. A. K., Zuo, X., Clark, P. K., Campbell, S. J., Wang, Y.-X., Rein, A. (2011). Solution Properties of Murine Leukemia Virus Gag Protein: Differences from HIV-1 Gag. *Journal of Virology*, Vol. 85, 23, 12733-12741.
- Deng, N., Li, Z., Pan, C., Duan, H. (2015). freeQuant: A Mass Spectrometry Label-Free Quantification Software Tool for Complex Proteome Analysis. *The Scientific World Journal*, 137076, 11.

- D'haeseleer, P. (2005). How does gene expression clustering work? *Nature Biotechnology*, Vol. 23, 12, 1499-1501.
- Eidhammer, I., Barsnes, H., Eide, G. E., Martens, L. 2013. *Computational and Statistical Methods for Protein Quantification by Mass Spectrometry*, p. 186-204. 1st ed. John Wiley & Sons, Ltd.
- Everitt, B. S., Landau, S., Leese, M., Stahl, D. 2011. *Cluster Analysis*, p. 71-110. 5th ed. John Wiley & Sons, Ltd.
- Gelfand, M. 2011. Dynamic programming. One algorithmic key for many biological locks, p. 66-92. *In* Pevzner, P., Shamir, R. (Eds.), *Bioinformatics for biologists*, Cambridge University Press.
- Gipe, G. W. (1976). Using residual analysis to search for specification errors. *Decision Sciences*, 7(1):40-56.
- Guo, C., Liu, S., Wang, J., Sun, M.-Z., Greenaway, F. T. (2012). ACTB in cancer, *Clinica Chimica Acta*, 417, 39-44.
- Habek, M. (2003). New approach to gene expression analysis. *DDT*, Vol. 8, 10.
- Hussain, S., Hazarika, G. (2011). Enhanced Hierarchical Clustering for Genome Databases. *IJCSI International Journal of Computer Science Issues*, Vol.8, 4, 1, 1694-0814.
- Jiang, D., Tang, C., Zhang, A. (2004). Cluster Analysis for Gene Expression Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, 11, 1370-1386.
- Kabacoff, R. I. 2011. *R in action. Data analysis and graphics with R*. Manning Publications Co.

- Kobiyama, K., Kawashima, A., Jounai, N., Takeshita, F., Ishii, K. J., Ito, T., Suzuki, K. (2013). Role of extrachromosomal histone H2B on recognition of DNA viruses and cell damage. *Front. Genet.* 4:91, doi: 10.3389/fgene.2013.00091.
- Kurg, R., Reinsalu, O., Jaguar, S., Õunap, K., Võsa, L., Kasvandik, S., Padari, K., Gildemann, K., Ustav, M. (2016). Biochemical and proteomic characterization of retrovirus Gag based microparticles carrying melanoma antigens. *Scientific Reports*, 6:29425, doi: 10.1038/srep29425.
- Lavallée-Adam, M., Rauniyar, N., McClatchy, D., B., Yates, J. R. (2014). PSEA-Quant: A Protein Set Enrichment Analysis on Label-Free and Label-Based Protein Quantification Data. *Journal of Proteome Research*, 13, 5496-5509.
- Liu, W., Wang, T., Chen, S. (2009). Hierarchical Clustering of Gene Expression Data with Divergence Measure. *IEEE*, 978-1-4244-2902-8/09.
- Luo, F., Tang, K., Khan, L. (2003). Hierarchical Clustering of Gene Expression Data. *Proceedings of the Third IEEE Symposium on Bioinformatics and BioEngineering (BIBE'03)*, 0-7695-1907-5/03.
- Mayer, M. P., Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6): 670-684.
- Megger, D. A., Bracht, T., Meyer, H. E., Sitek, B. (2013). Label-free quantification in clinical proteomics. *Biochimica et Biophysica Acta*, 1834, 1581–1590.
- Miles, J. 2014. R Squared, Adjusted R Squared. *Wiley StatsRef: Statistics Reference Online*. University of York, York, UK.
- Murtagh, F., Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *WIREs Data Mining and Knowledge Discovery*, 2:86-97.

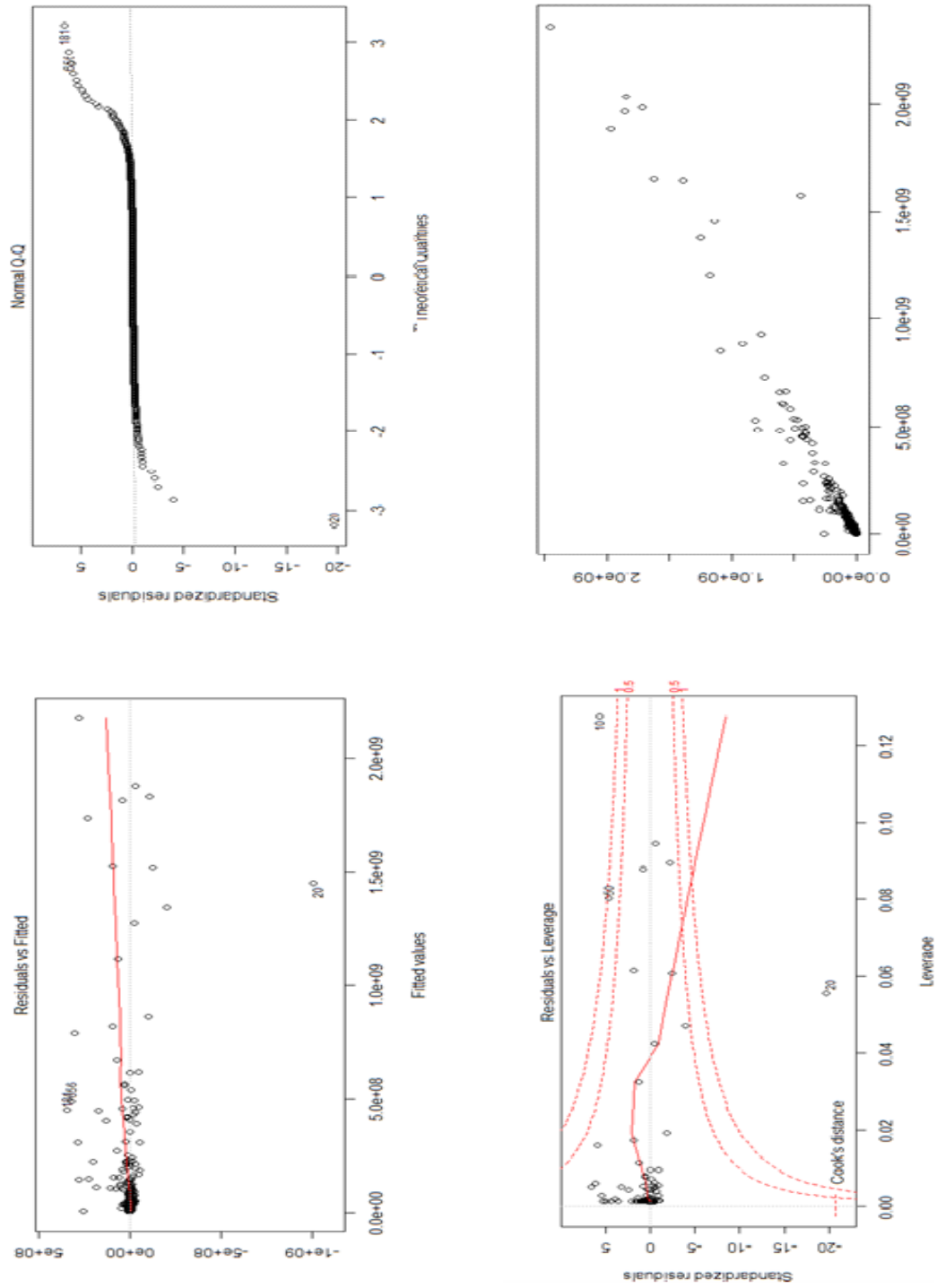
- Nahnsen, S., Bielow, C., Reinert, K., Kohlbacher, O. (2013). Tools for Label-free Peptide Quantification. *Molecular & Cellular Proteomics*, 12(3), 549–556.
- Pocsfalvi, G., Stanly, C., Vilasi, A., Fiume, I., Capasso, G., Turiak, L., Buzas, E., Vekey, K. (2015). Mass spectrometry of extracellular vesicles. *Mass Spectrometry Reviews*, 2016, 35, 3–21.
- Priya, R., Dass, J. F. P., Siva, R. (2015). Gene Expression Prediction and Hierarchical Clustering Analysis of Plant CCD genes. *Plant Mol Biol Rep*, doi: 10.1007/s11105-015-0950-2.
- Pujalto, P. B. 2011. Study of gene expression representation with Treelets and hierarchical clustering algorithms, p. 13-28. *Enginyeria de Telecomunicacio, Departament de Teoria del Senyal i Comunicacions, Grup de Processament d'Imatge i Video*.
- Ramos, D., Pellin-Carcelen, A., Agusti, J., Murgui, A., Jorda, E., Pellin, A., Monteagudo, C. (2015). Deregulation of Glyceraldehyde-3-Phosphate Dehydrogenase Expression During Tumor Progression of Human Cutaneous Melanoma. *Anticancer Research* 35: 439-444.
- Seal, S., Komarina, S., Aluru, S. (2004). An optimal hierarchical clustering algorithm for gene expression data. *Information Processing Letters* 93, 143-147.
- Seo, J., Shneiderman, B. (2003). Understanding Hierarchical clustering Results by Interactive Exploration of Dendrograms: A Case Study with Genomic Microarray Data. *Technical Reports from UMIACS*, 02/2003, 35.
- Shiga, M., Mamitsuka, H. (2011). Clustering genes with expression and beyond. *WIREs Data Mining and Knowledge Discovery*, 1:496-511.
- Sirbu, A., Bocicor, M. I. (2013). A Dynamic Approach for Hierarchical Clustering of Gene Expression Data. *IEEE*, 978-1-4799-1494-4/13.

- Smorenburg, S. M., Griffini, P., Tiggelman, A. M. B. C., Moorman, A. F. M., Boers, W., Van Noorden, J. F. (1996). A2-Macroglobulin Is Mainly Produced by Cancer Cells and Not by Hepatocytes in Rats With Colon Carcinoma Metastases in Liver. *Hepatology*, Vol. 23, 3, 560-570.
- Szeto, L. K., Liew, A. W.-C., Yan, H., Tang, S. (2006). Gene Expression Data Clustering and Visualization Based on a Binary Hierarchical Clustering Framework. *Conferences in Research and Practice in Information Technology*, Vol. 19.
- Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y. (2004). Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell*, Vol. 116, 1, 51-61.
- Tasoulis, S. K., Plagianakos, V. P., Tasoulis, D. K. 2010. Projection Based Clustering of Gene Expression Data, p. 228-239. *In* Masulli, F., Peterson, L., Tagliaferri, R. (Eds.), *CIBB 2009, LNBI 6160*, Springer-Verlag Berlin Heidelberg.
- Zhang, Q., Zhang, Y. (2006). Hierarchical clustering of gene expression profiles with graphics hardware acceleration. *Pattern Recognition Letters*, 27, 676-681.
- Zhu, W., Smith, J. W., Huang, C.-M. (2010). Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology*, 2010:840518.
- Wang, M., You, J.-S., Bemis, K. G., Fitzpatrick, P. G. 2008. Label-free Mass Spectrometry-Based Protein Quantification Technologies in Protein Biomarker Discovery, p. 211-230. *In* Wang, F. (Ed.), *Methods in Pharmacology and Toxicology: Biomarker Methods in Drug Discovery and Development*, Humana Press, Totowa, NJ.
- Wang, M., You, J., Bemis, K. G., Tegeler, T. J., Brown, D. P. G. (2008). Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in Functional Genomics*, 7(5):329–339.

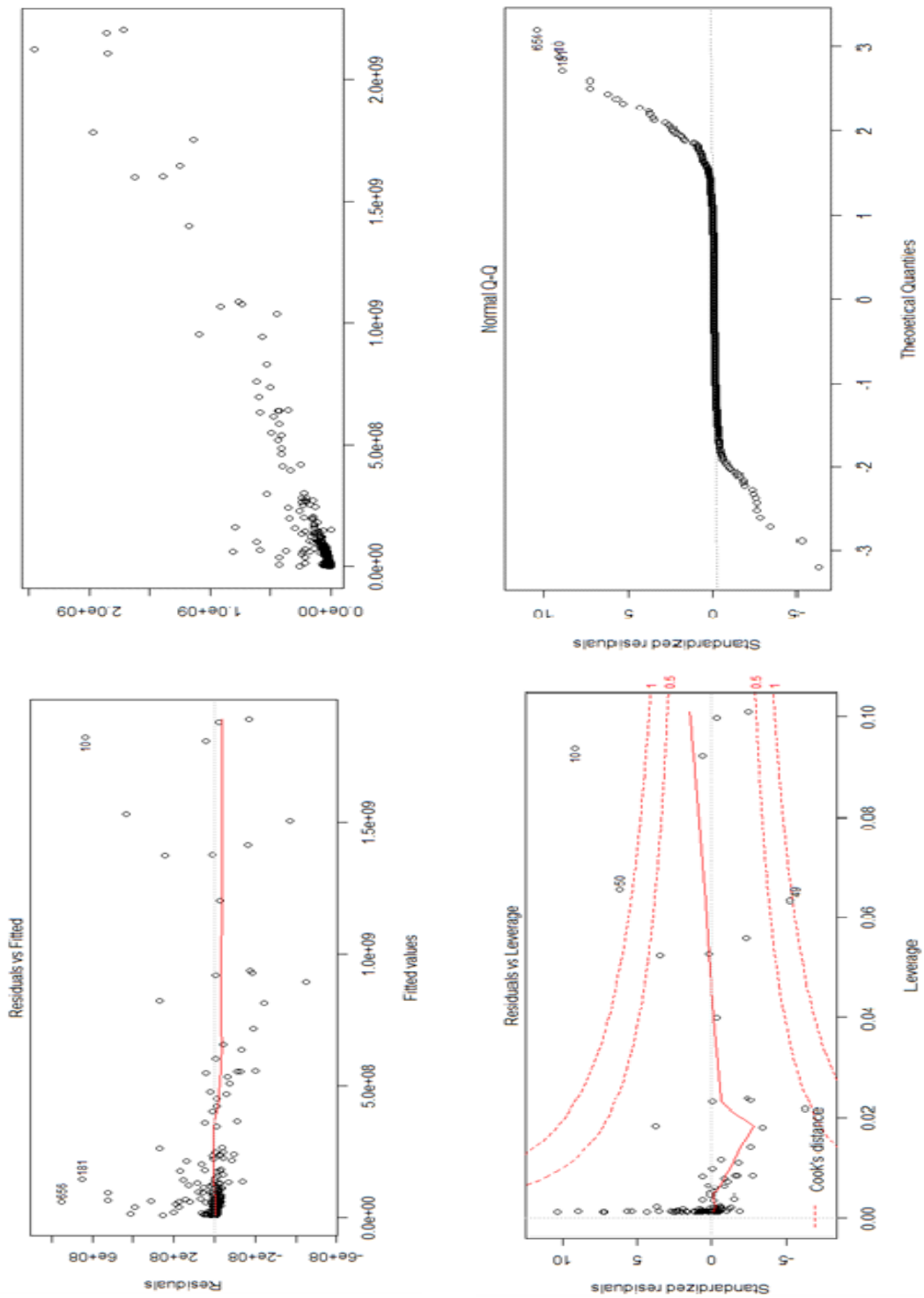
Kasutatud veebiaadressid

<https://data.library.virginia.edu/diagnostic-plots/>

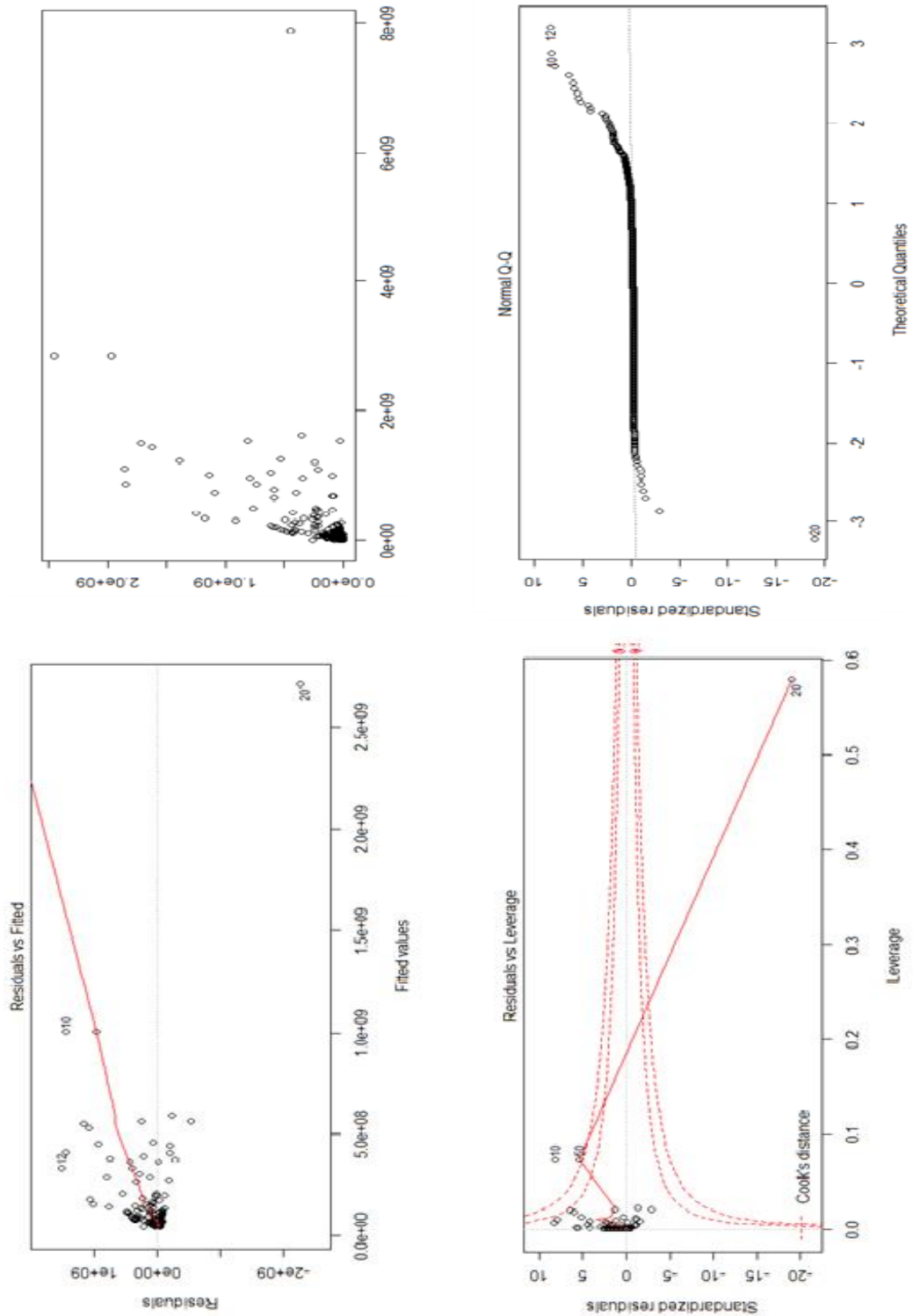
Lisa 1



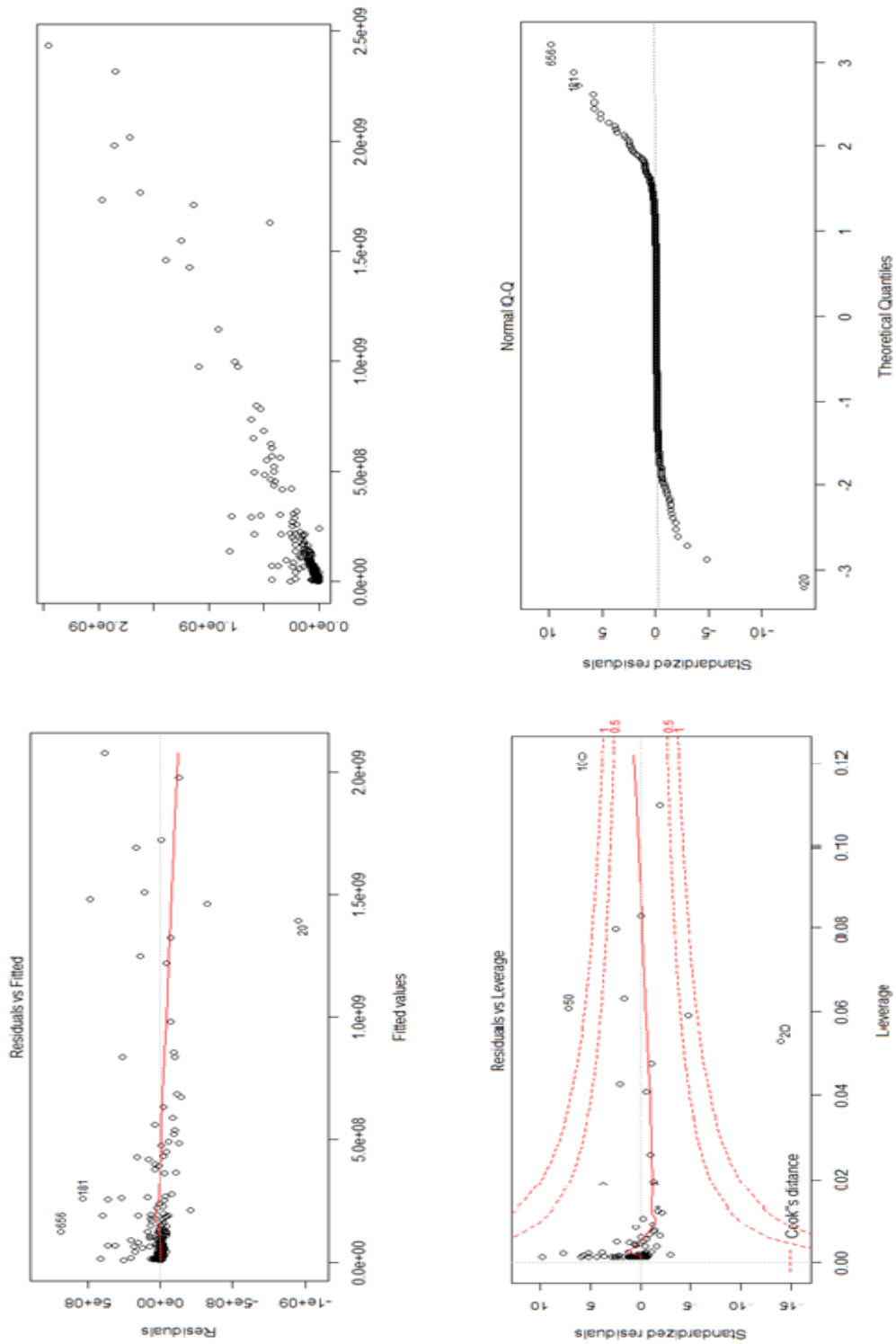
Joonis 5. MAGEA4 vs MART1.



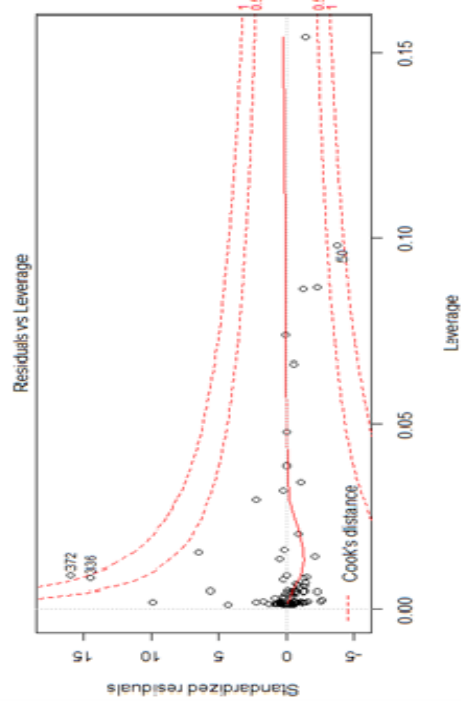
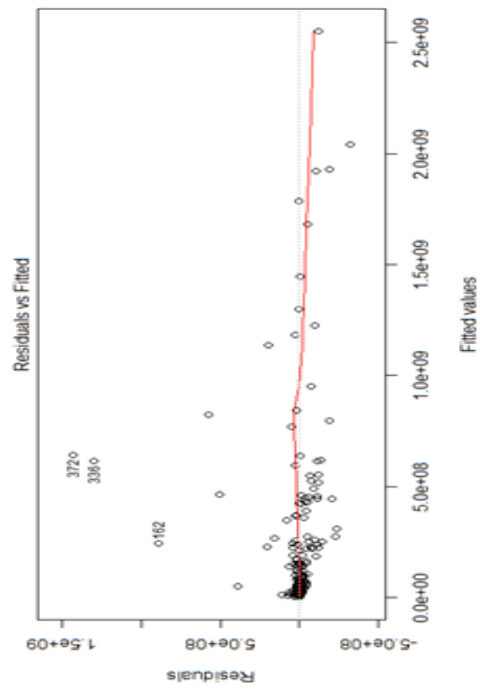
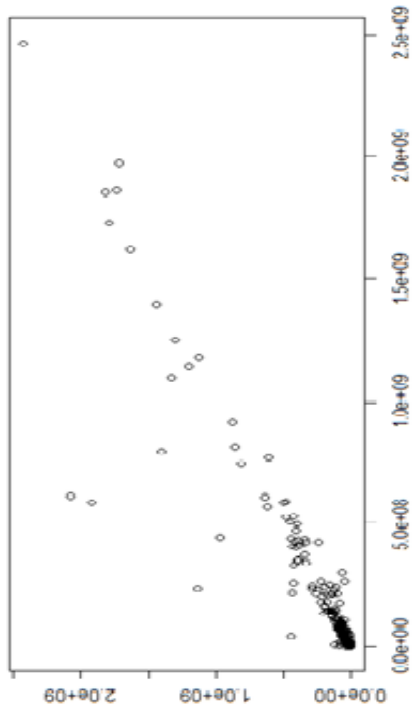
Joonis 6. MAGEA4 vs MCAM.



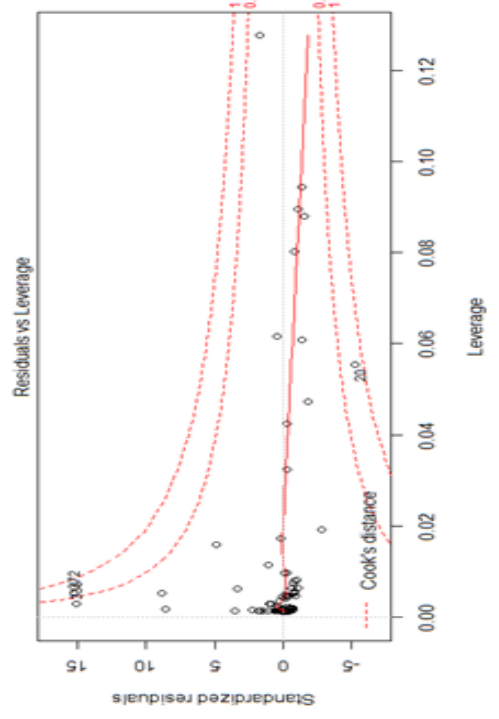
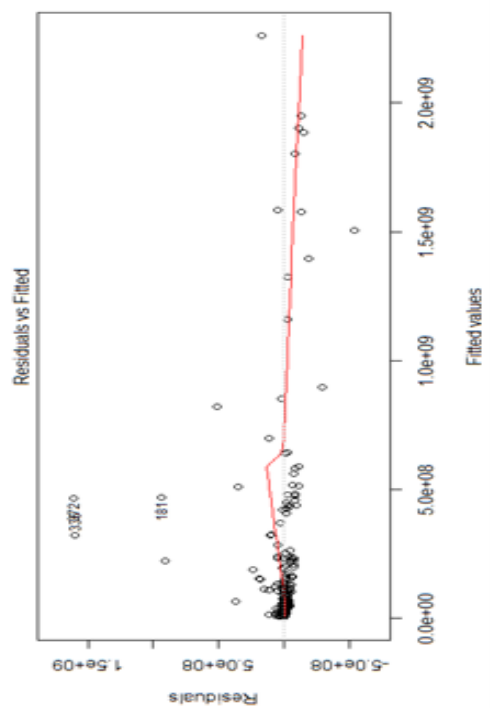
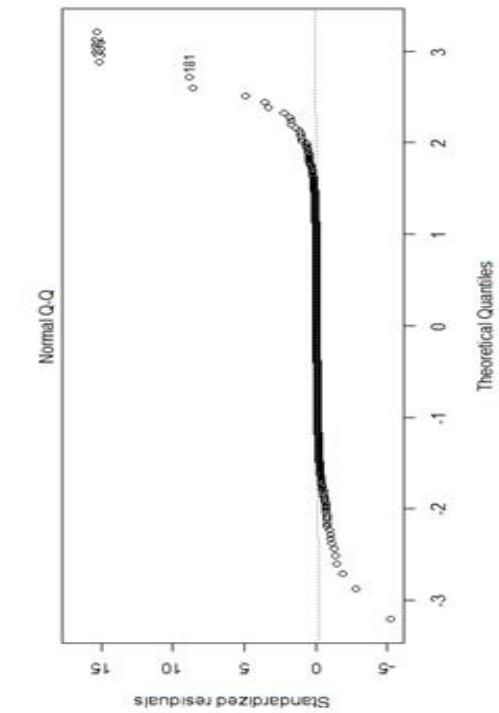
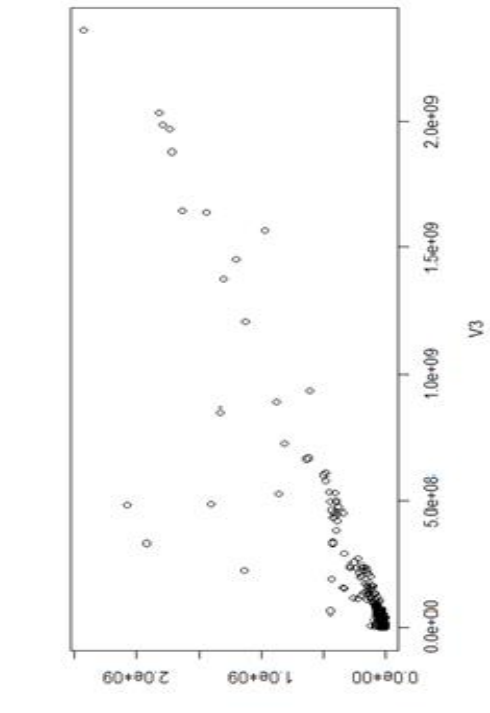
Joonis 7. MAGEA4 vs MLV gag.



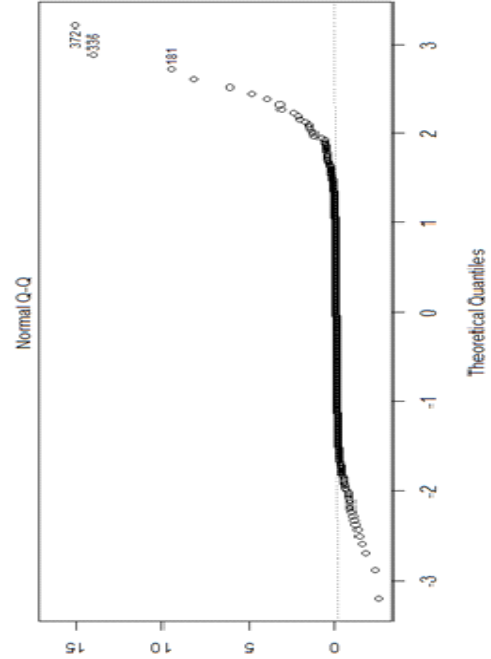
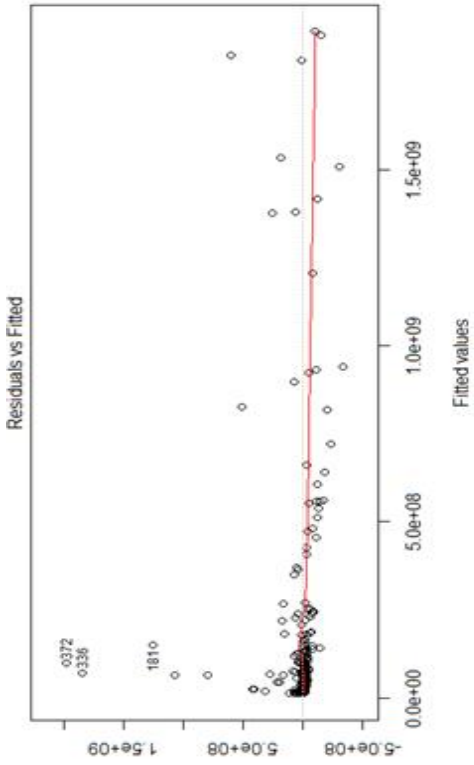
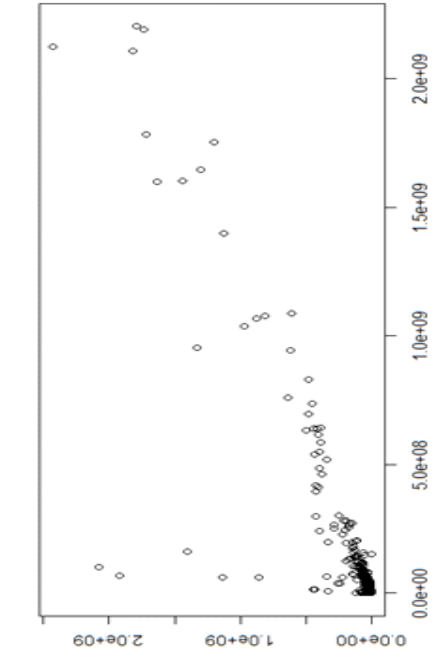
Joonis 8. MAGEA4 vs TRP1.



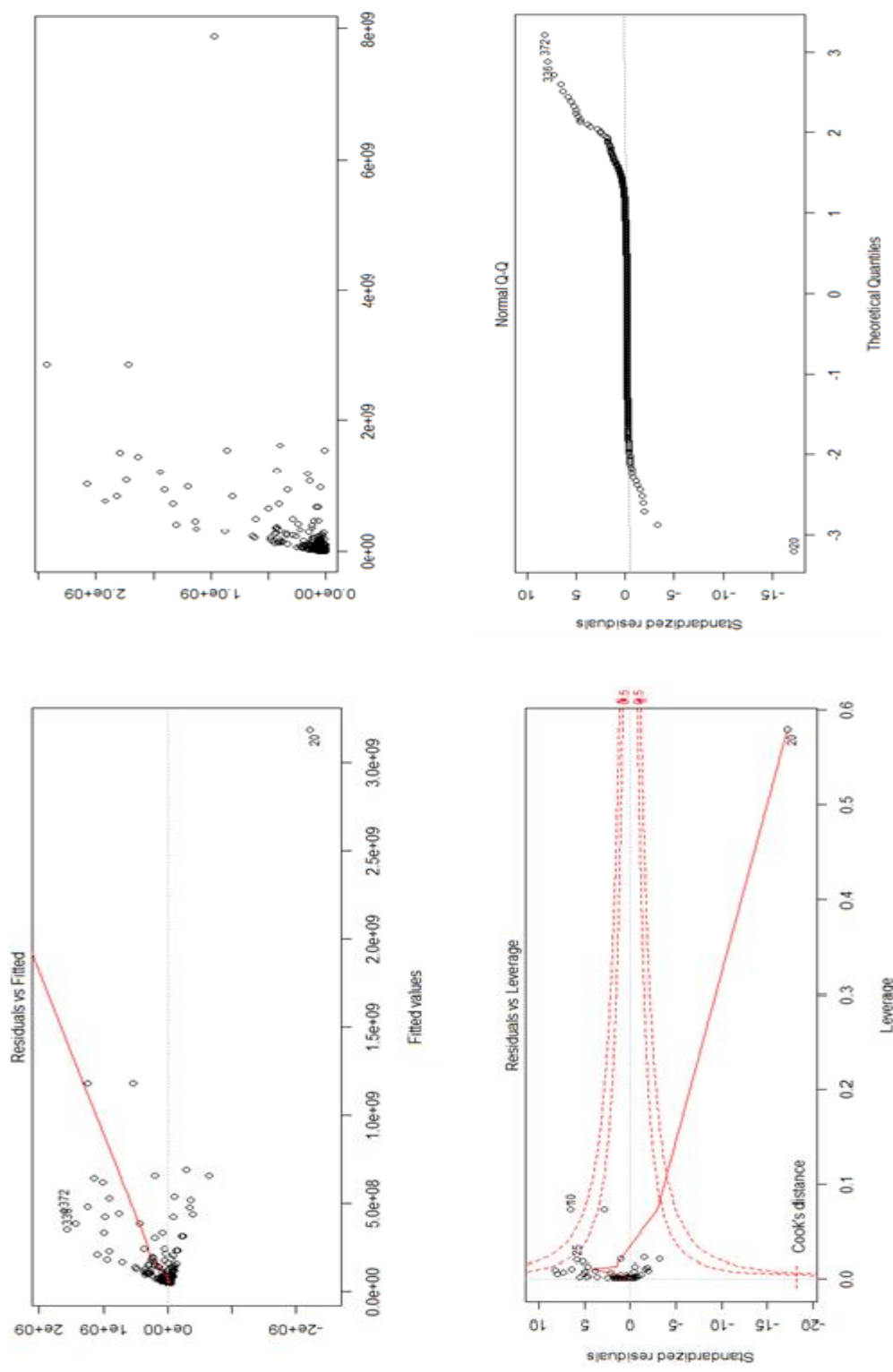
Joonis 9. MAGEA10 vs MAGEA4.



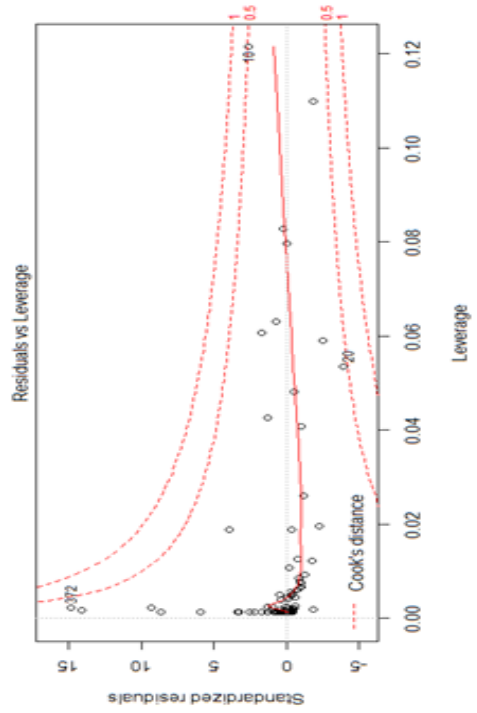
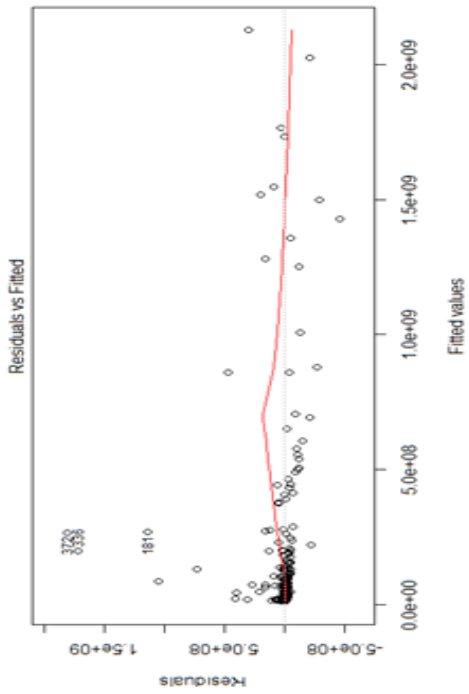
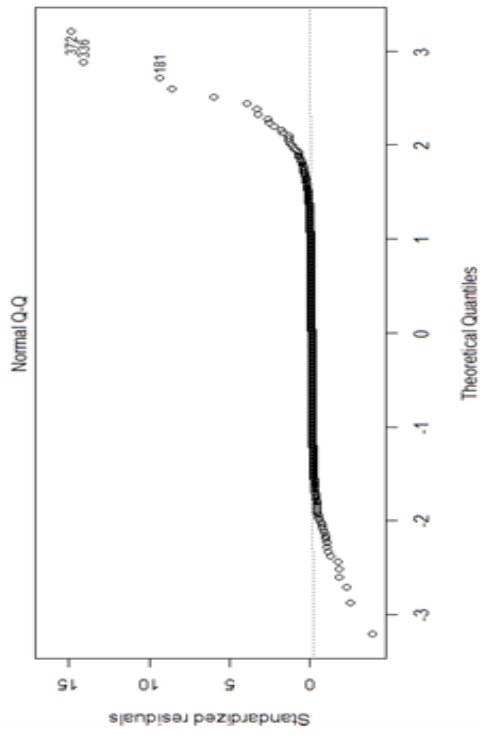
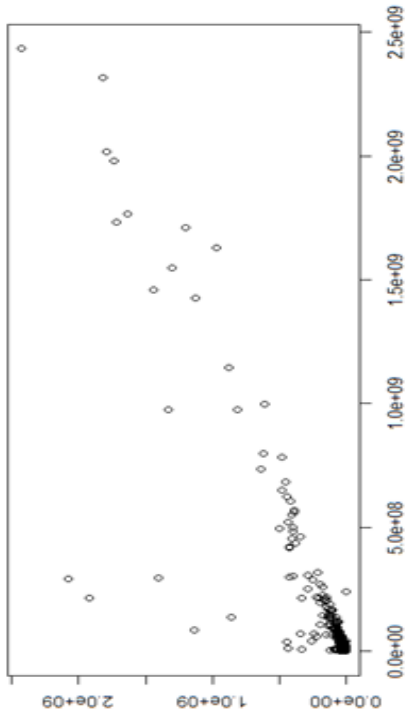
Joonis 10. MAGEA10 vs MART1.



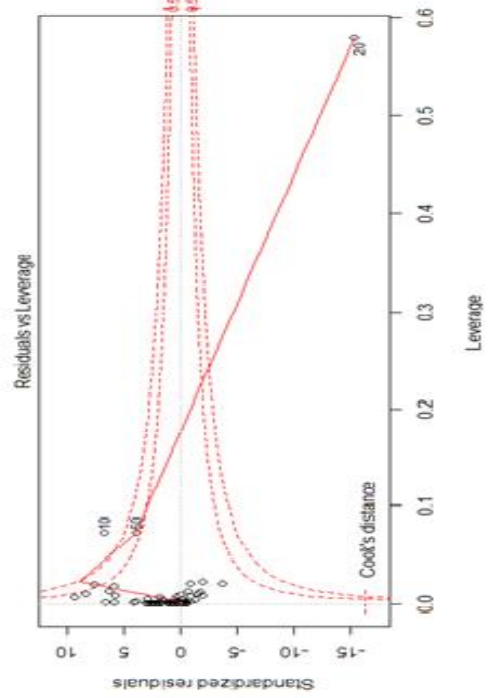
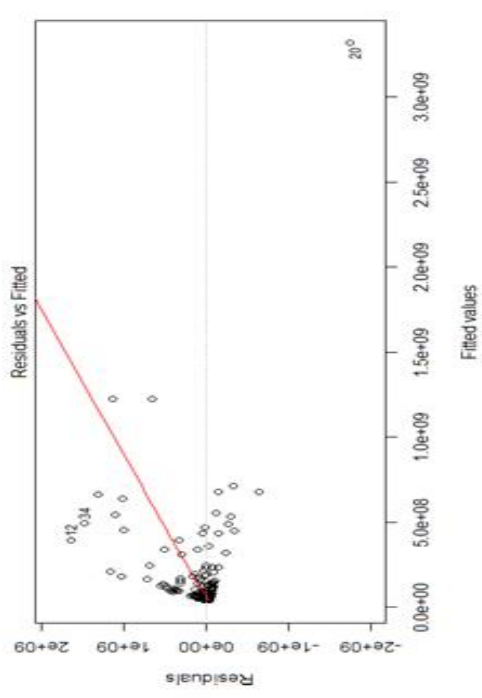
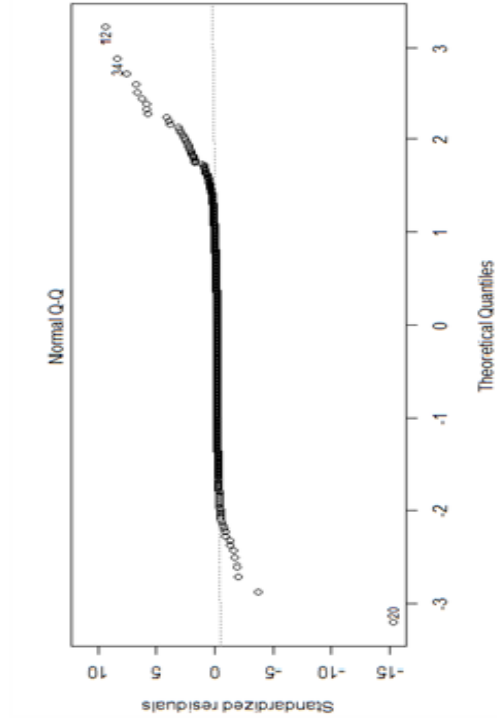
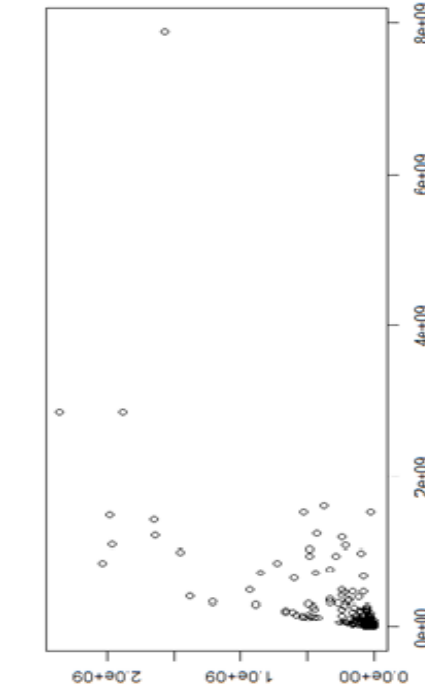
Joonis 11. MAGEA10 vs MCAM.



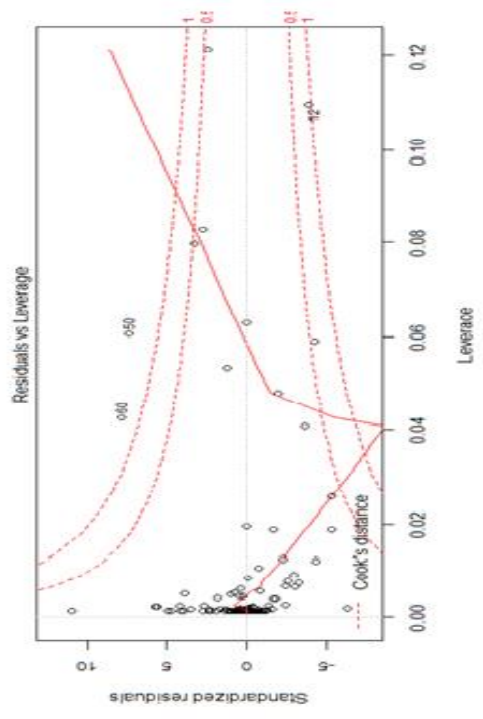
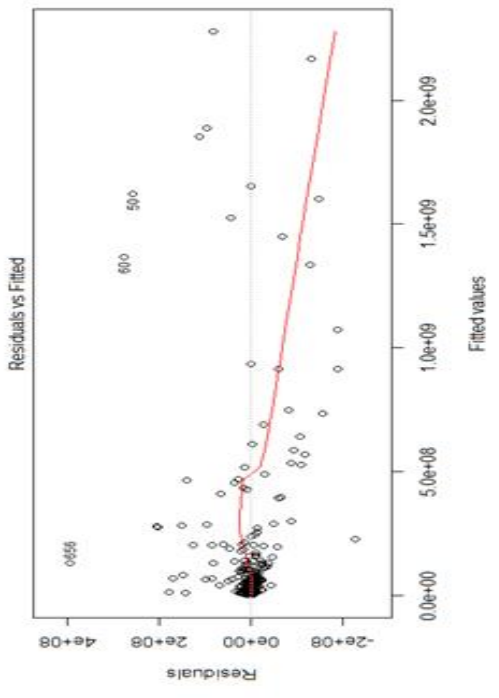
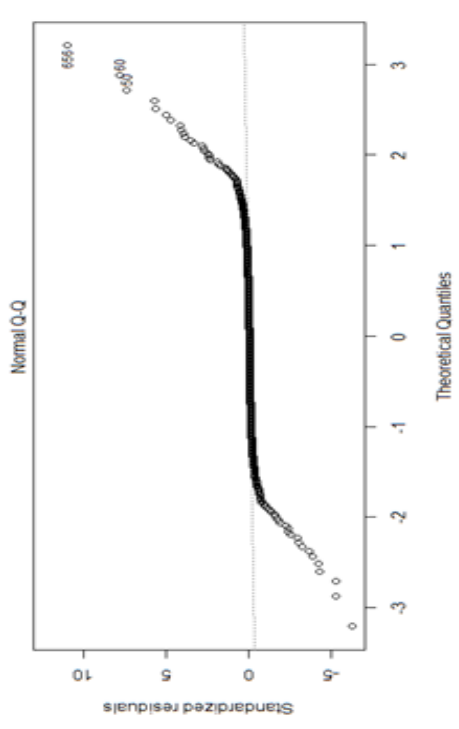
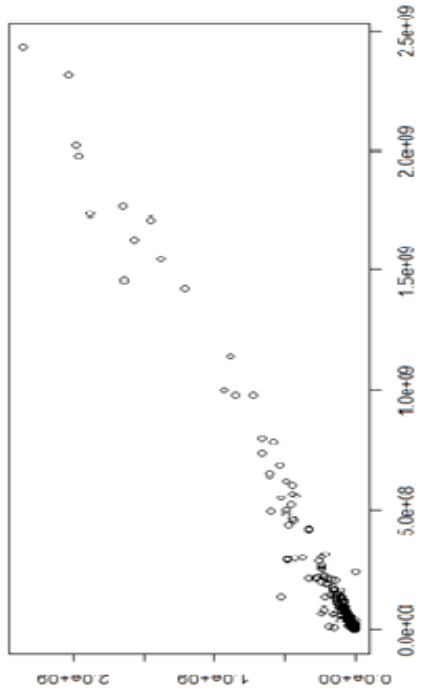
Joonis 12. MAGEA10 vs MLV gag.



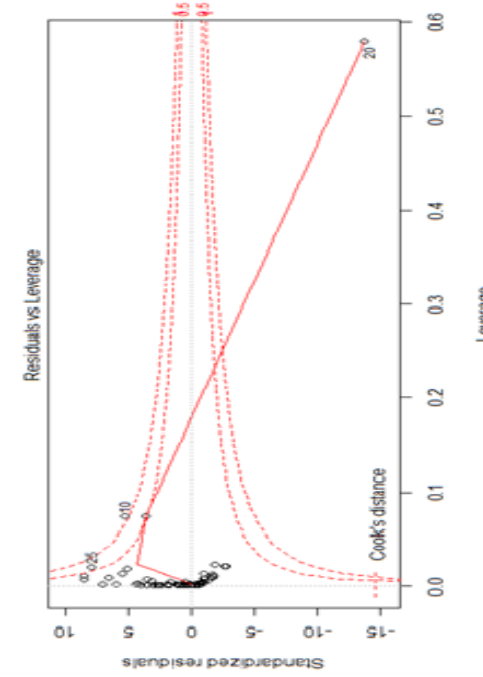
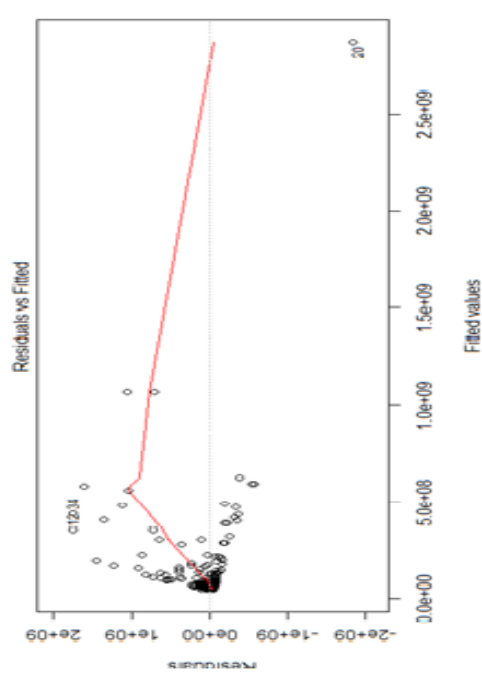
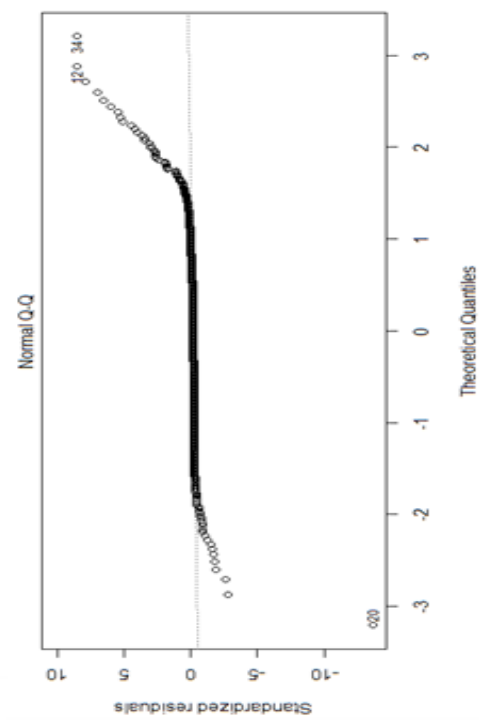
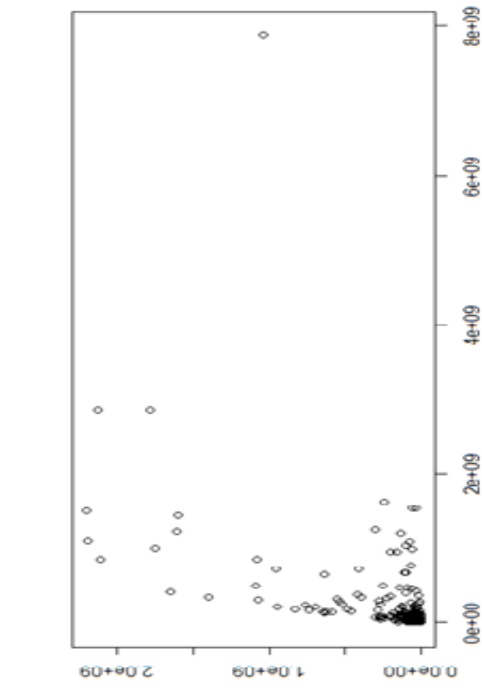
Joonis 13. MAGEA10 vs TRP1.



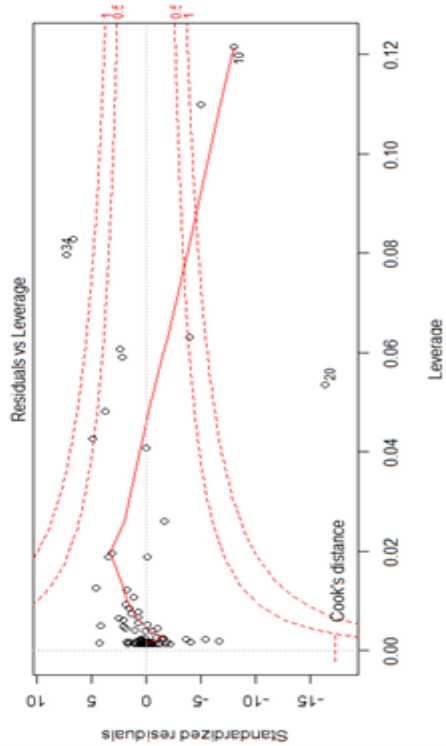
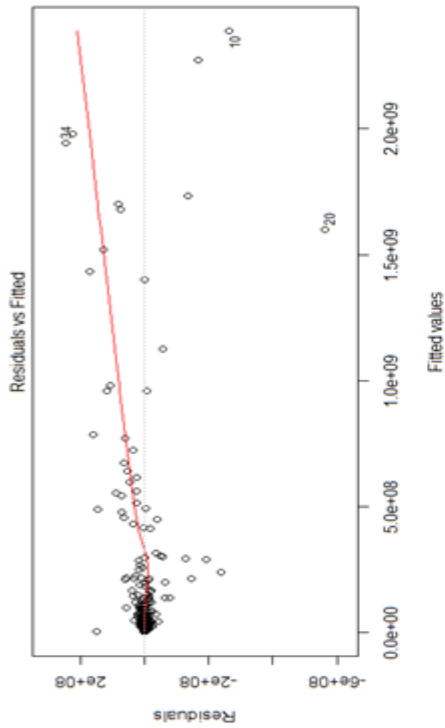
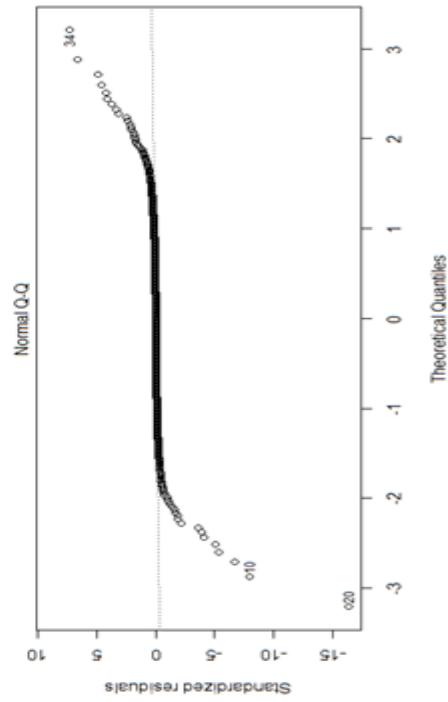
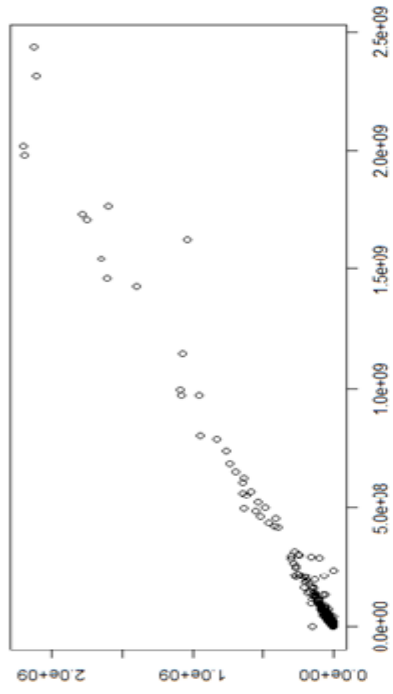
Joonis 15. MART1 vs MLV gag.



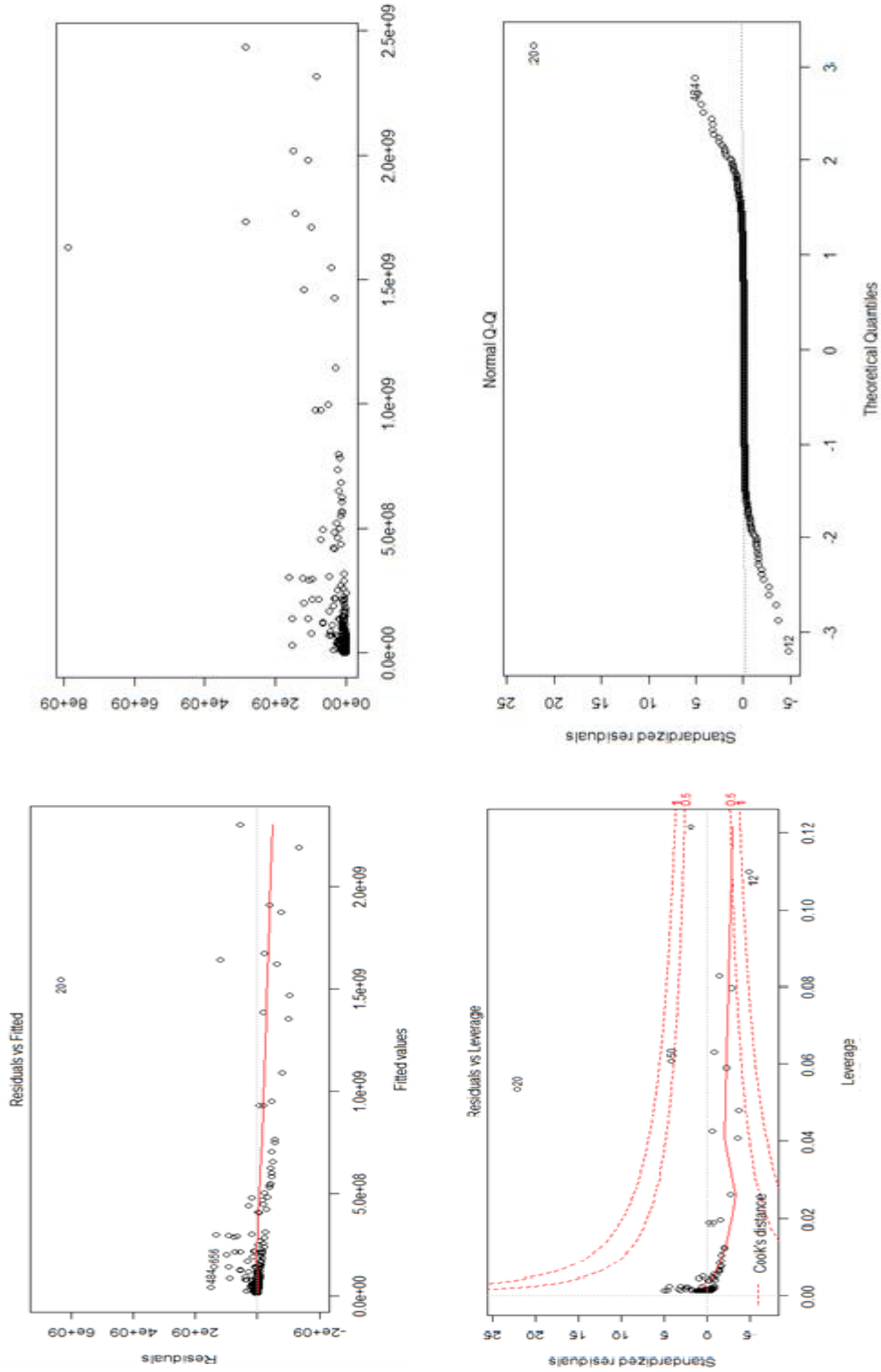
Joonis 16. MART1 vs TRP1.



Joonis 17. MCAM vs MLV gag.

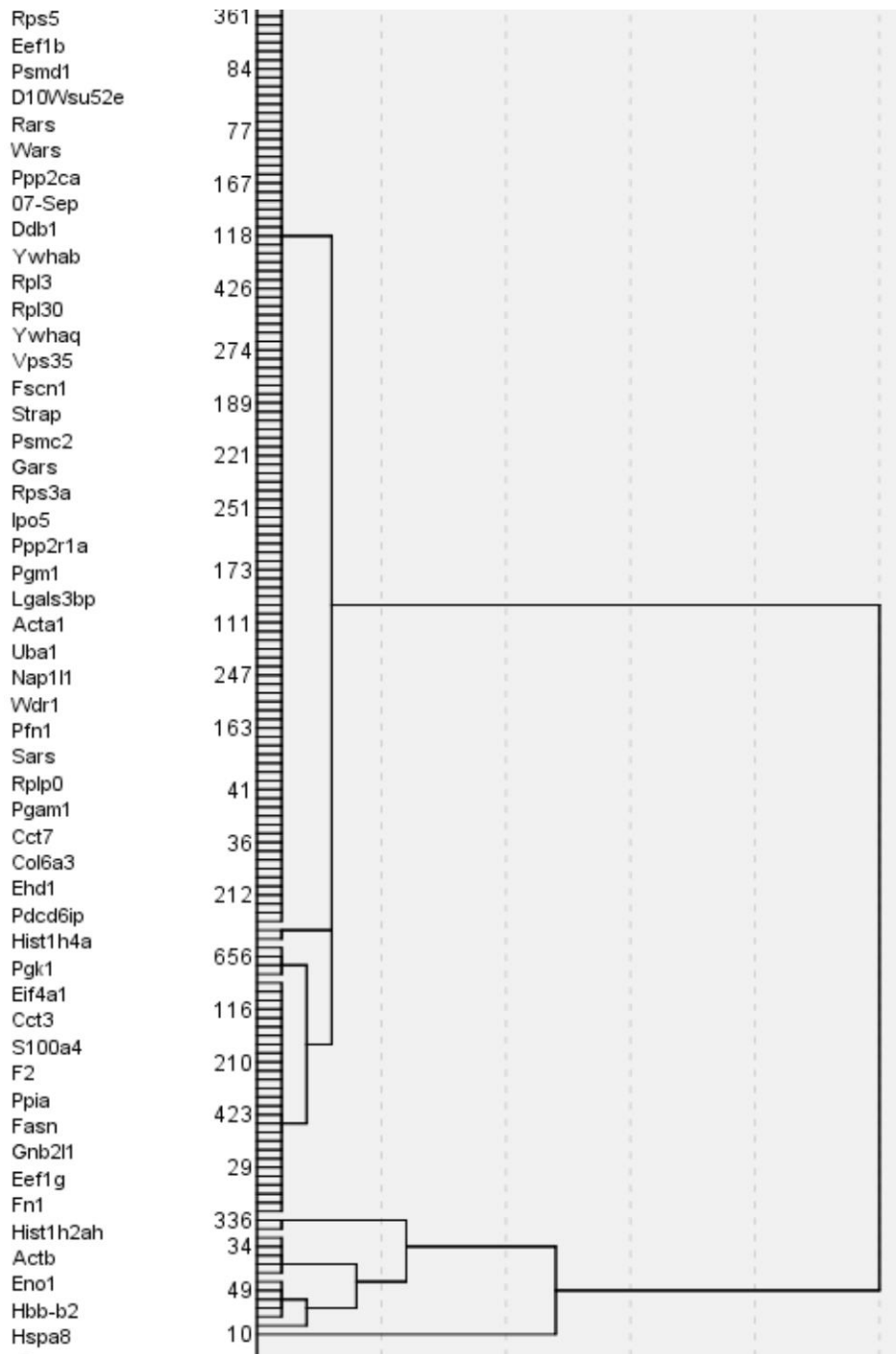


Joonis 18. MCAM vs TRP1.

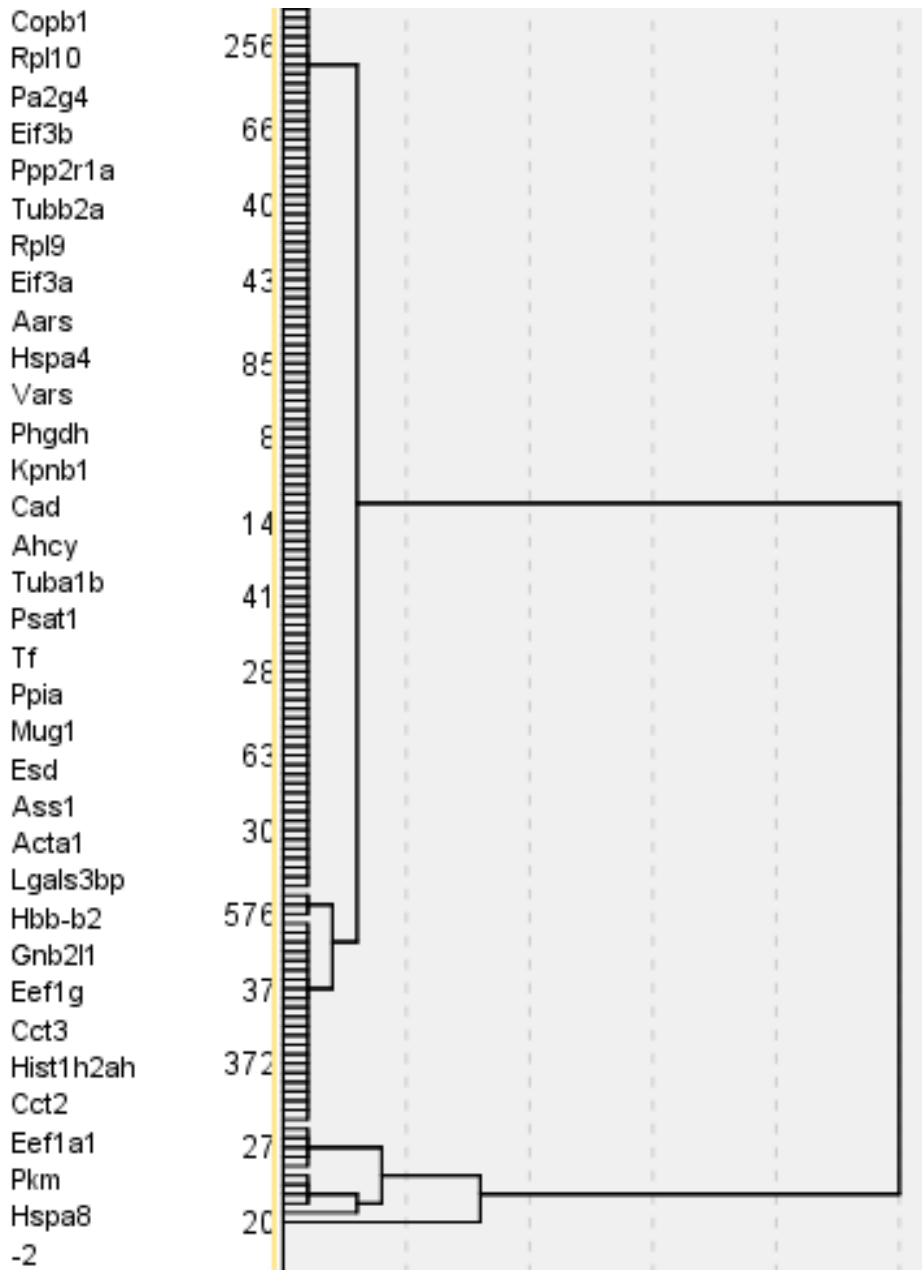


Joonis 19. MLV gag vs TRP1.

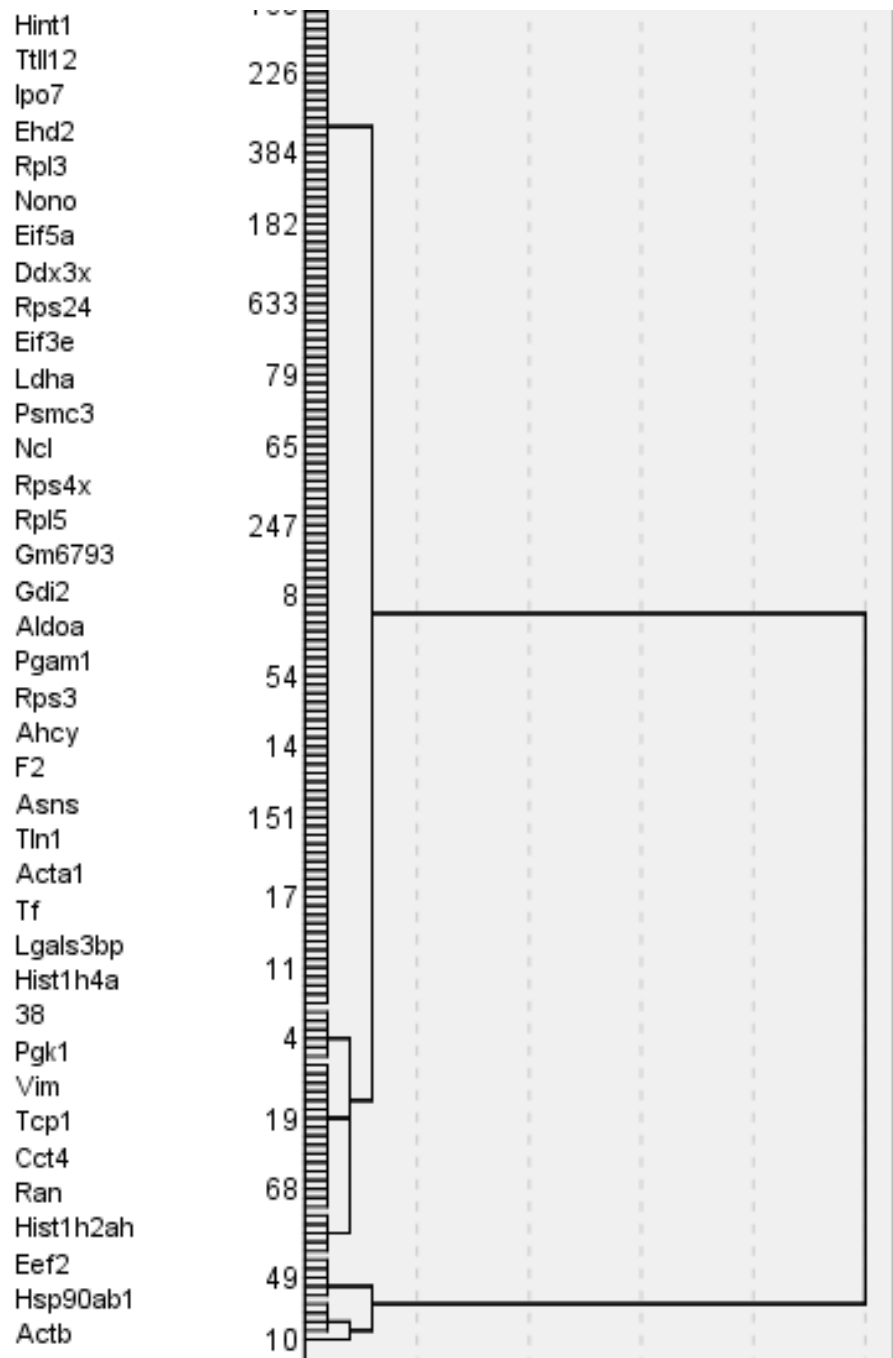
Lisa 2



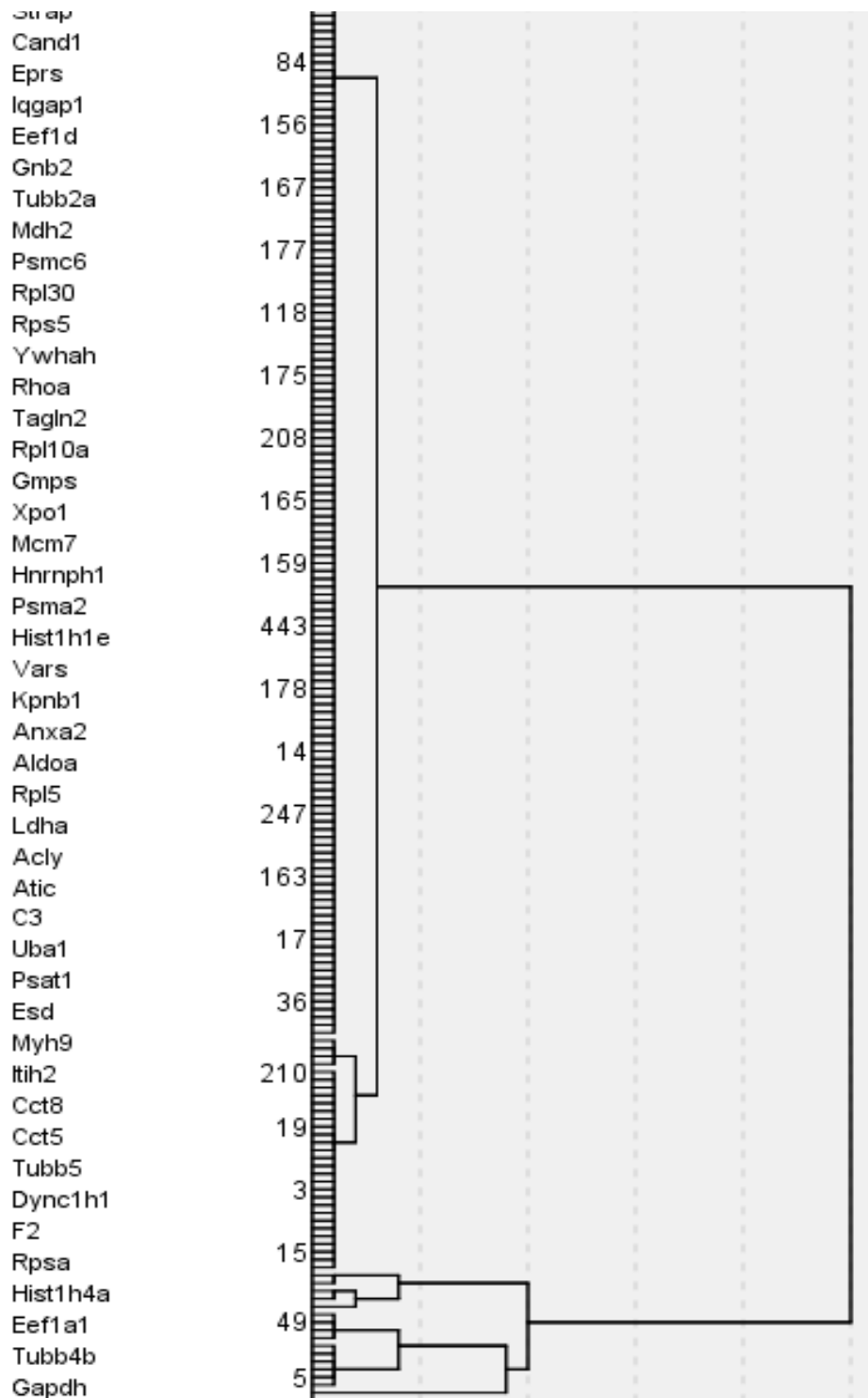
Joonis 20. MAGEA4 ja MAGEA10 dendrogramm.



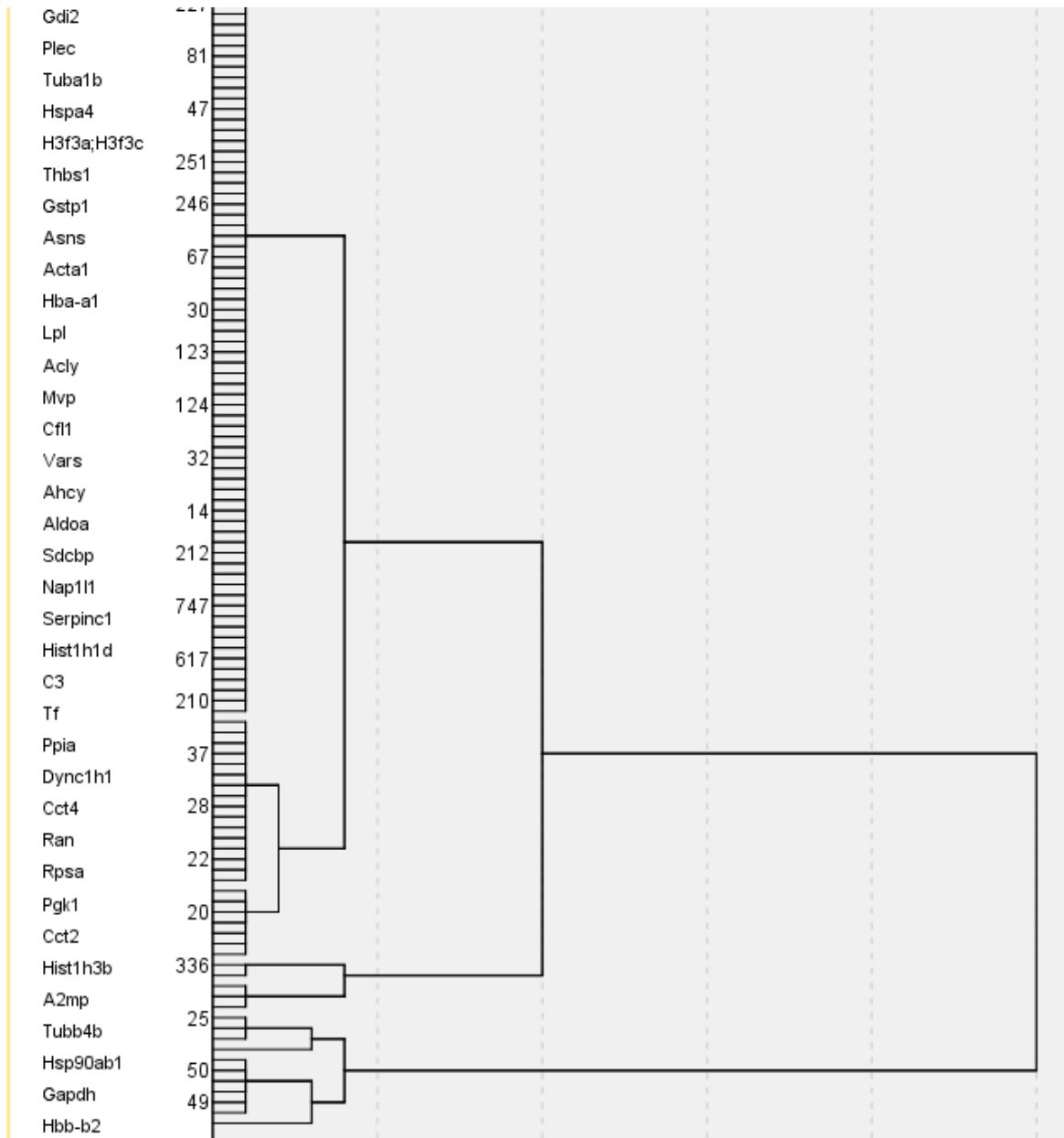
Joonis 21. MAGEA4 ja MART1 dendrogramm.



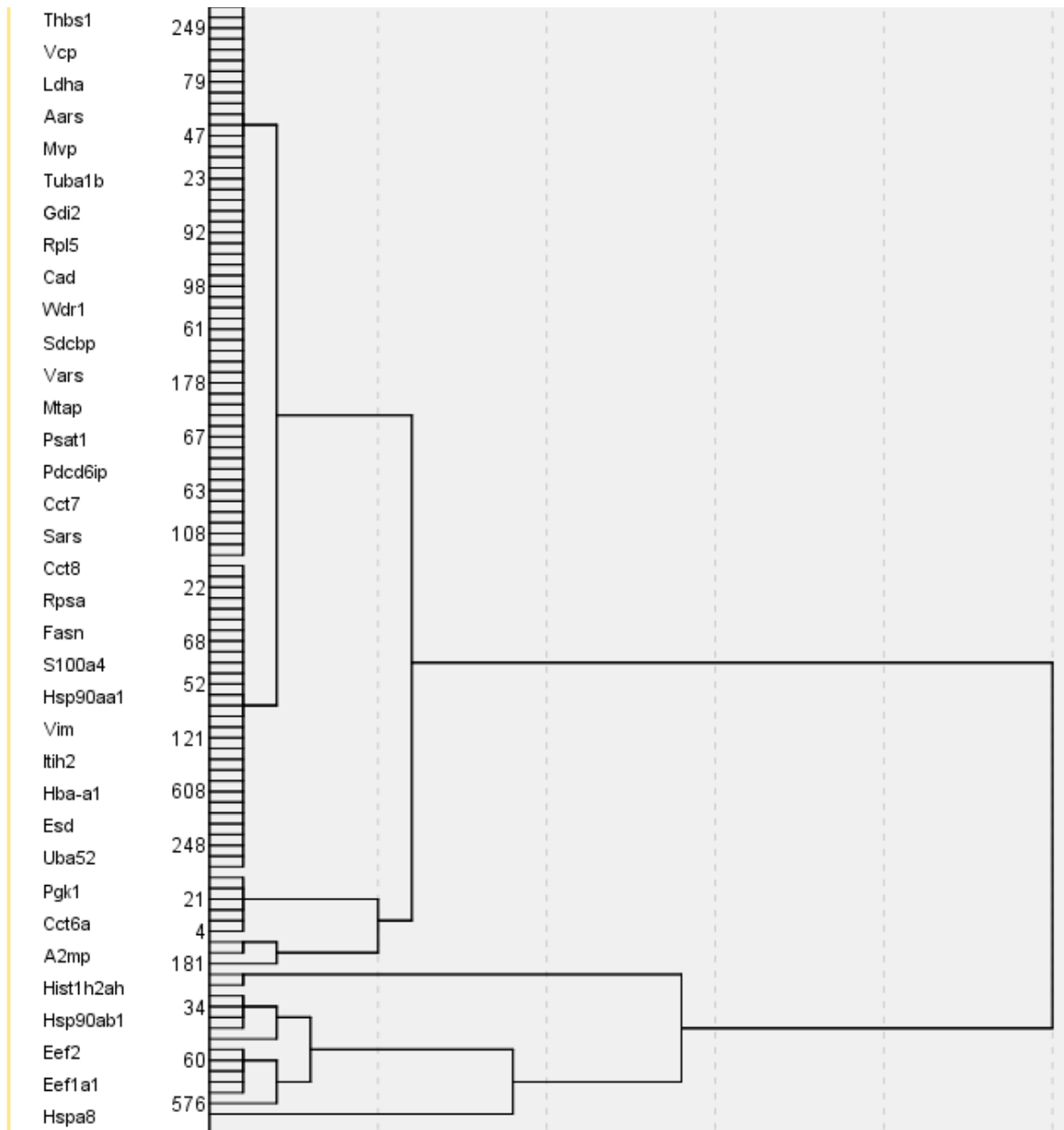
Joonis 22. MAGEA4 ja MCAM dendrogramm.



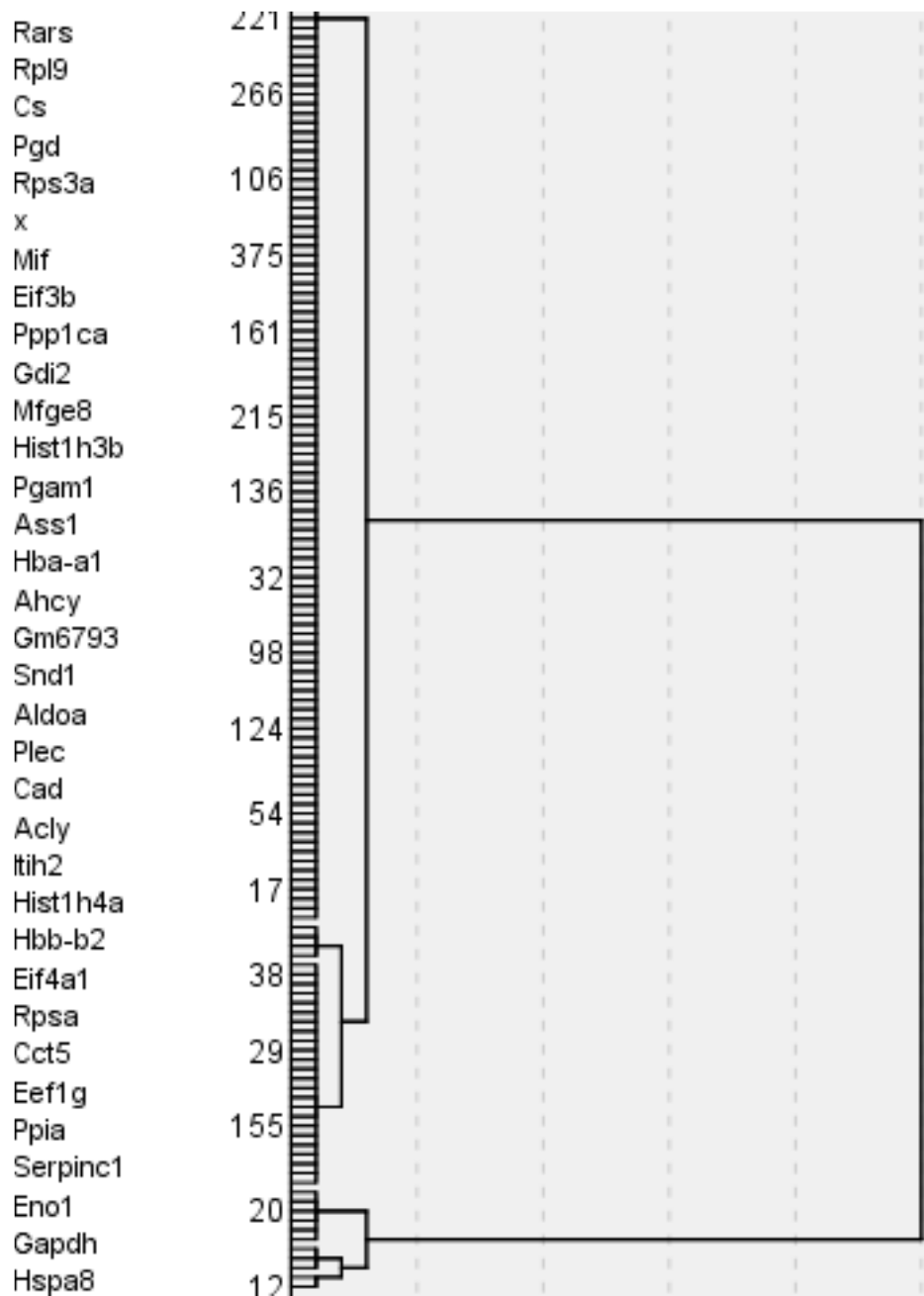
Joonis 23. MAGEA10 ja MART1 dendrogramm.



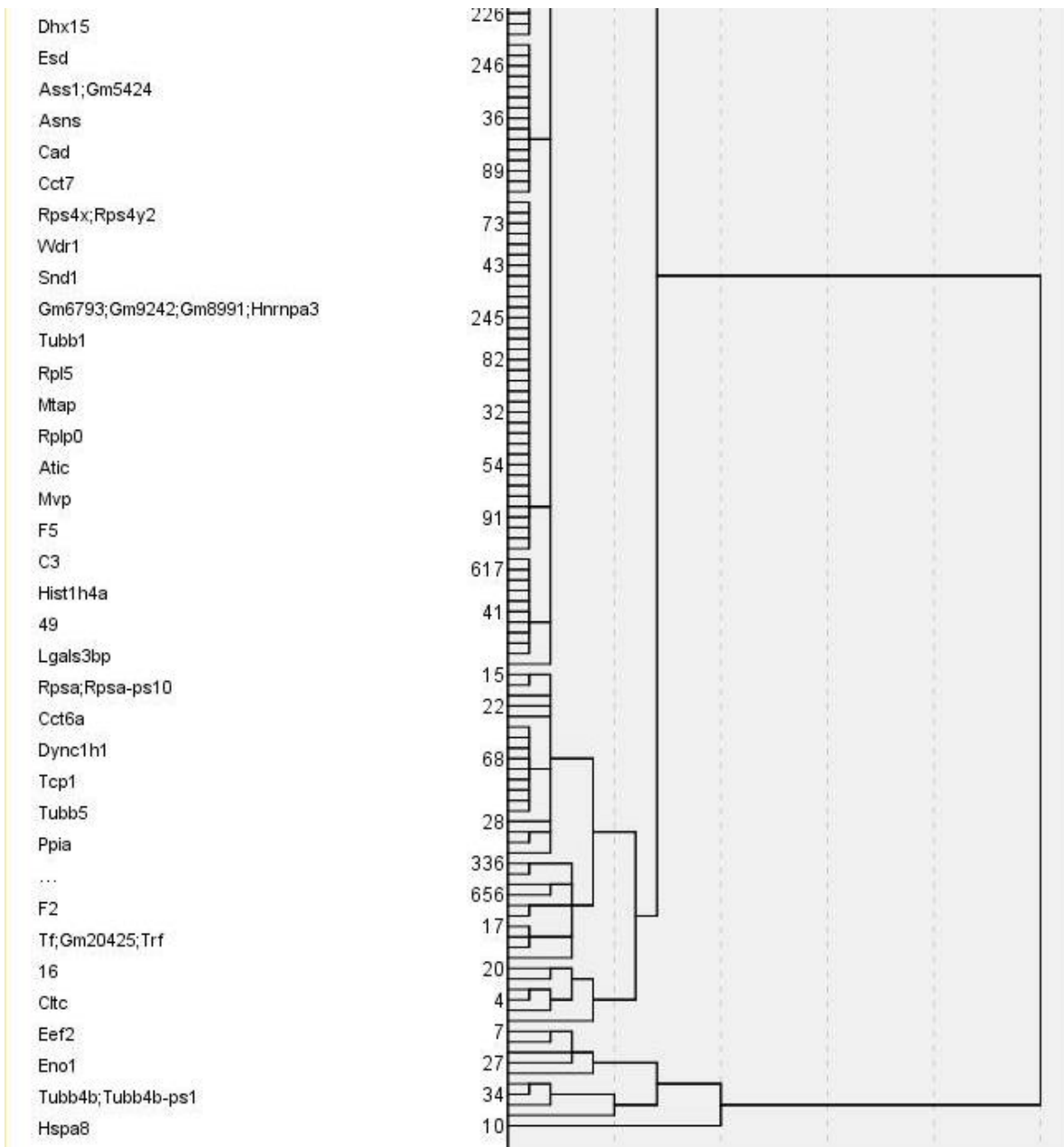
Joonis 24. MAGEA10 ja MCAM dendrogramm.



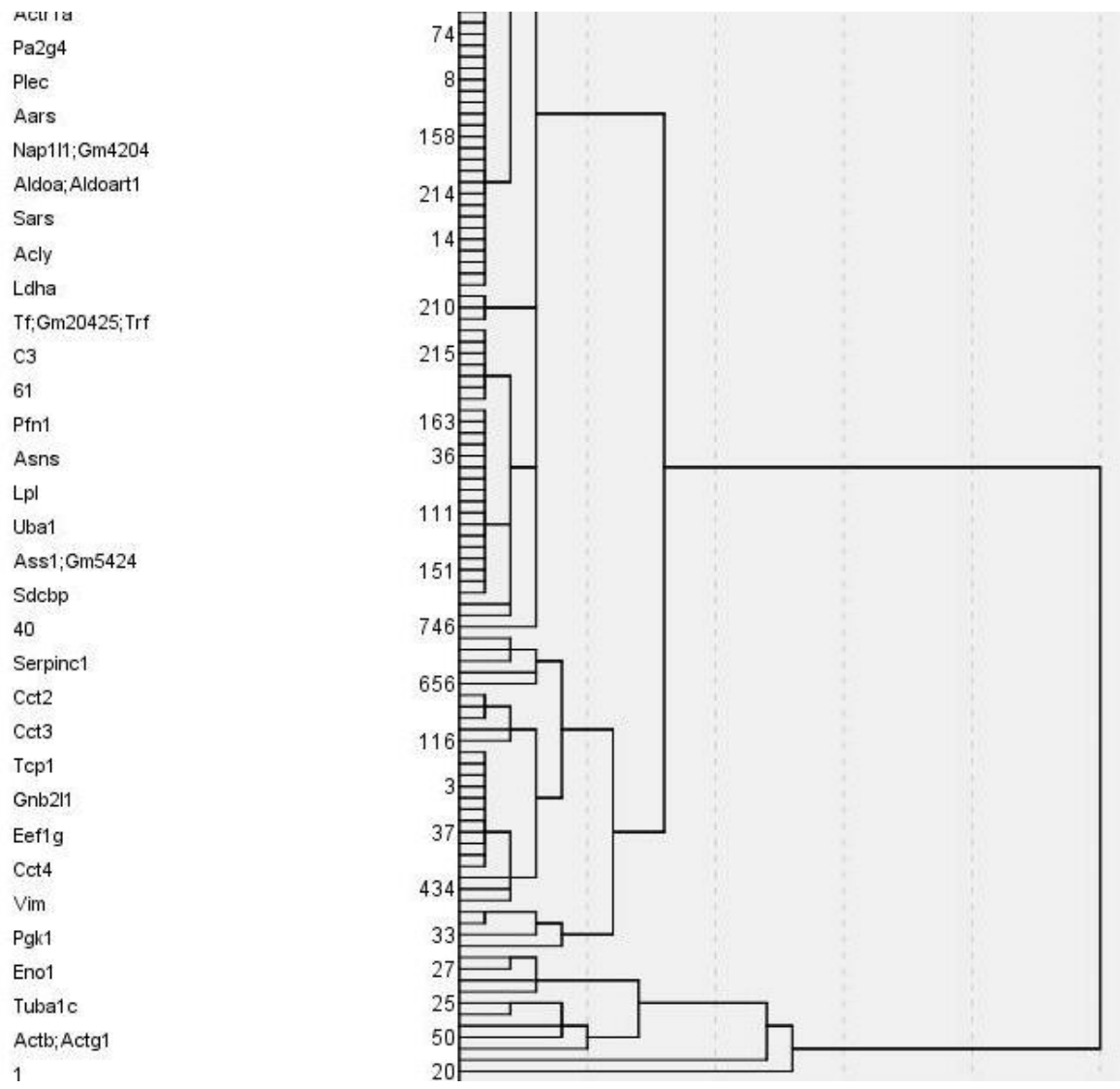
Joonis 25. MAGEA10 ja TRP1 dendrogramm.



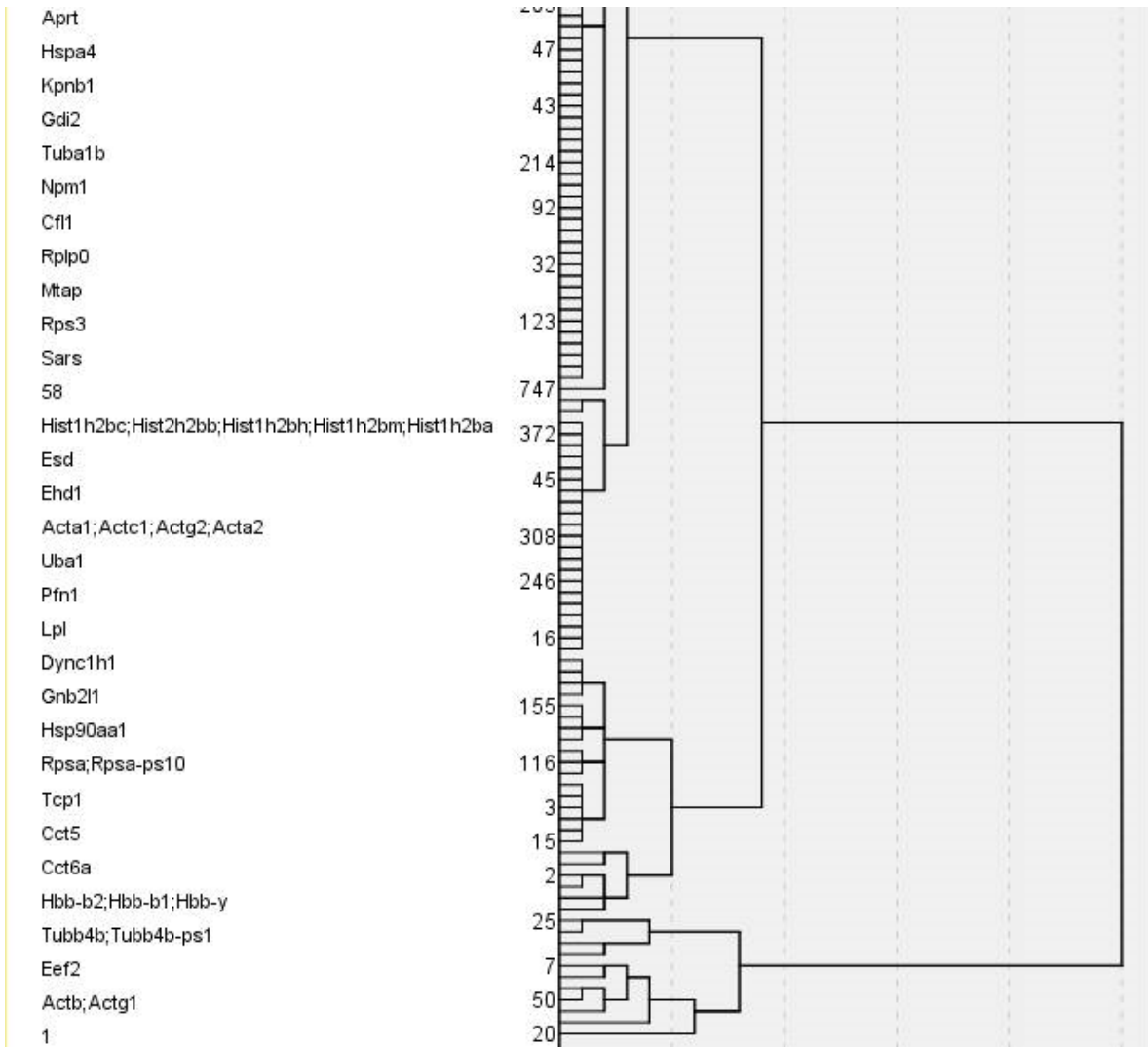
Joonis 26. MART1 ja TRP1 dendrogramm.



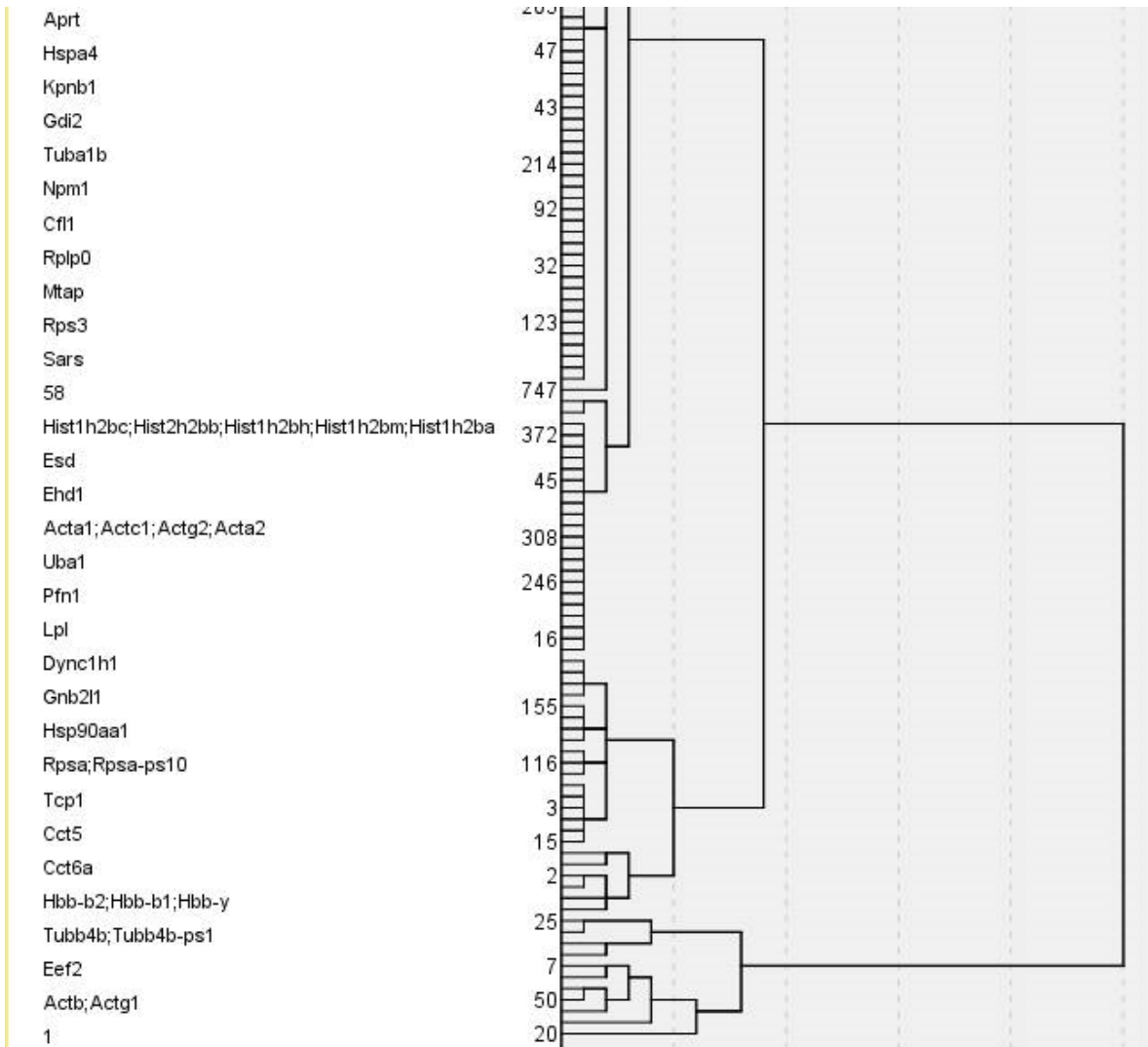
Joonis 27. MCAM ja MART1 dendrogramm.



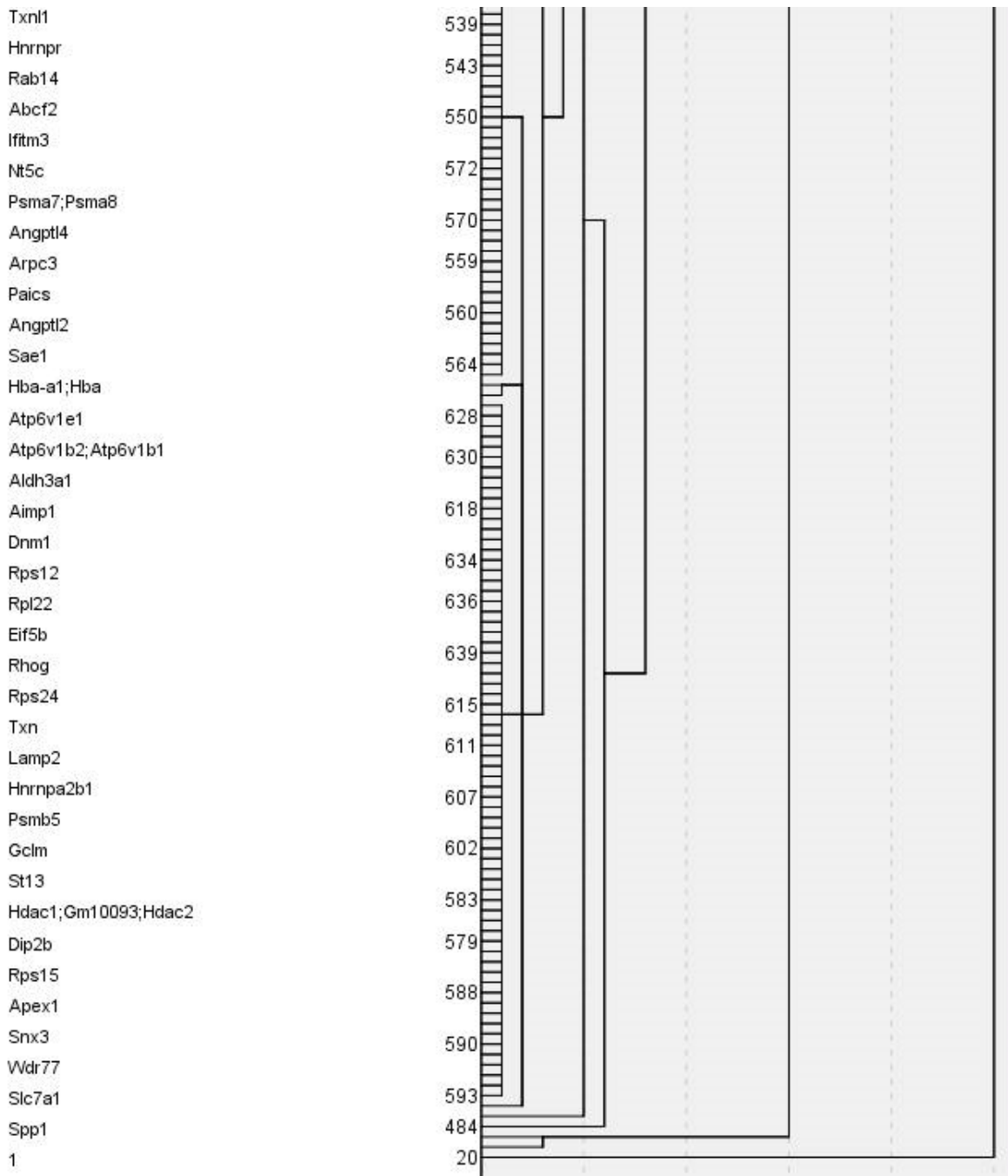
Joonis 28. MAGEA4 ja TRP1 dendrogramm.



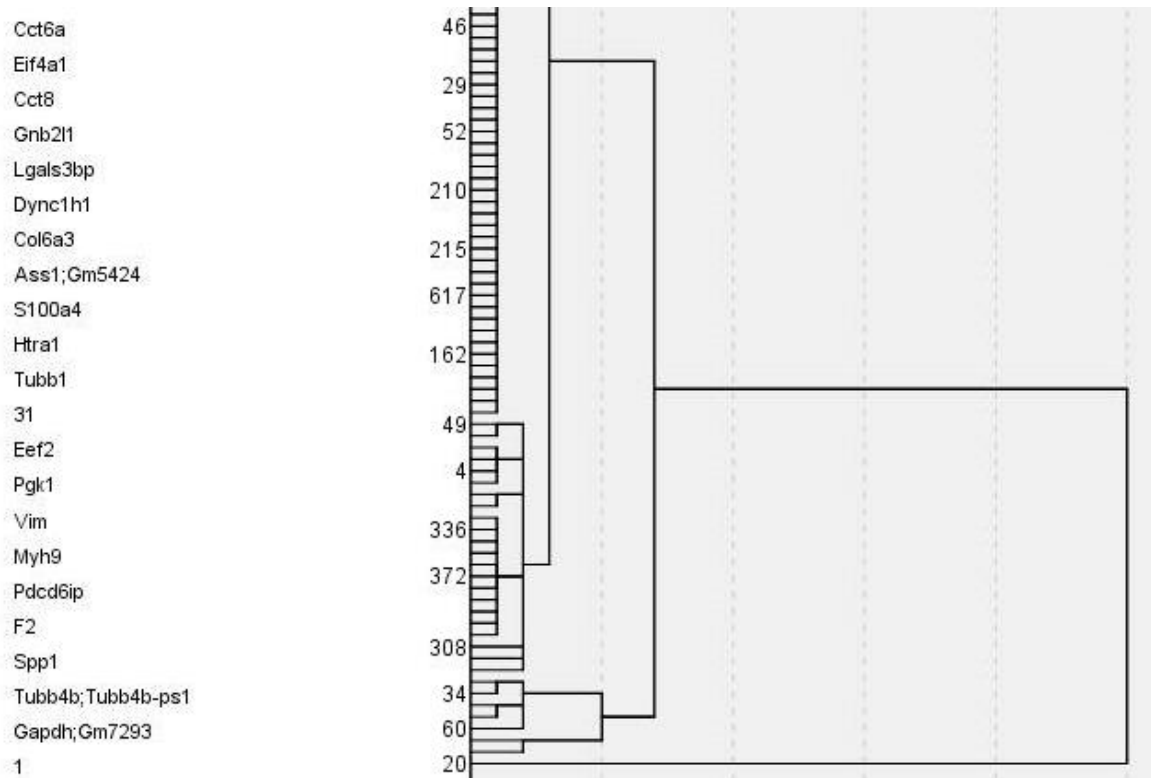
Joonis 29. MCAM ja TRP1 dendrogramm.



Joonis 30. MLV gag ja TRP1 dendrogramm.



Joonis 31. MLV gag ja MAGEA10 dendrogramm.



Joonis 32. MLV gag ja MAGEA4 dendrogramm.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina,

Ranel Tammela _____,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Melanoomi antigeene kandvate viiruslaadsete partiklite proteoomianalüüs _____

_____,

(lõputöö pealkiri)

mille juhendaja on

Reet Kurg _____,

(juhendaja nimi)

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **22.08.2016**